Statistika i osnovna mjerenja

Uvod

M. Makek

2017/2018

Pravila izvođenja kolegija

- Nastava:
  - 2 + 3 sata (predavanja + vježbe) kroz cijeli semestar
  - Laboratorijske vježbe - svaki student radi 3 vježbe
- Kolokviji:
  - 1. kolokvij – uvjet za pohađanje laboratorijskih vježbi
  - 2. i 3. kolokvij – nastavno gradivo
  - Uz određeni broj bodova iz kolokvija moguće je oslobađanje od pismenog ispita (pravila na web stranici)
- Uvjet za potpis i pristupanje ispitu:
  - položen prvi kolokvij i pozitivno ocijenjene sve tri laboratorijske vježbe

Prvi kolokvij

15.3.2018.

u 10h (termin redovite nastave)

Gradivo:

- Račun pogreške
- Grafički prikaz rezultata
- Metoda najmanjih kvadrata

Kontakt i informacije

- Sva pravila izvođenja kolegija i dodatne informacije dostupne su na web stranici:

  http://www.phy.pmf.unizg.hr/~makek/som/

- Konzultacije:

  Mihael Makek

  Utorkom 13-14h, soba 326

Statistika i osnovna mjerenja

Račun pogreške

M. Makek

2017/2018

Obrada rezultata mjerenja

- Prvi dio kolegija – obrada rezultata mjerenja je uvod koji je nužan za pristup laboratorijskim vježbama
- Praktični pristup – primjeri, formule i zadatci
- Sadržaj:
  - ❑ **Račun pogreške**
  - ❑ **Grafički prikaz rezultata**
  - ❑ **Metoda najmanjih kvadrata**
- Cilj: statistički obraditi i pravilno prikazati rezultate mjerenja

RAČUN POGREŠKE

Vrste pogrešaka pri mjerenjima

Neovisna mjerenja

Ovisna mjerenja

Opća srednja vrijednost i nepouzdanost

Mjerenje fizikalnih veličina

- Cilj mjerenja je utvrditi brojčanu vrijednost neke veličine
- Zbog raznih utjecaja rezultat mjerenja $\boldsymbol{x_i}$ odstupa od prave vrijednosti veličine $\boldsymbol{X}$
- Odstupanje pojedine izmjerene vrijednosti od prave vrijednosti se naziva pogreškom mjerenja: $\boldsymbol{\Delta x = x_i - X}$
- Cilj eksperimenta je da izmjerena vrijednost bude što bliže pravoj vrijednosti i da pogreška mjerenja bude *pravilno procijenjena* kako bi rezultat bio valjan

Pogreške mjerenja

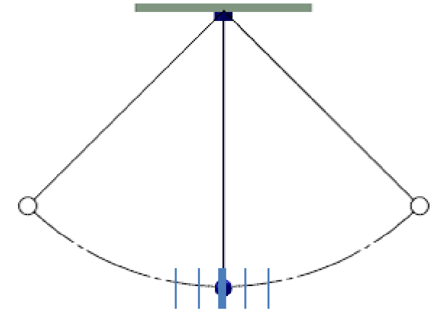
Razlikujemo tri tipa pogrešaka mjerenja:

- Slučajne pogreške
- Sistematske pogreške
- Grube pogreške

Grube pogreške

- Mogu nastati naglim poremećajem u okolini ili uređaju – npr. kvar
- Također mogu nastati zbog propusta eksperimentatora – npr. krivo očitanje ili zapisivanje rezultata
- Mjerenja za koja je ustanovljeno da sadrže grube pogreške treba odbaciti iz daljnje obrade ili ponoviti u ispravnim uvjetima

Sistematske pogreške (I)



Primjer:

- Mjerimo period titranja matematičkog njihala pomoću štoperice
- Moguće situacije:
  a) Vrijeme na štoperici odstupa od stvarnog vremena za konačan iznos. Npr. kad štoperica pokazuje 60 s, u stvarnosti je to 59 s

     → sva mjerenja vremena su sustavno manja za 59/60 ~ 1.7%

  b) Prilikom pritiska na gumb štoperice potrebno je neko vrijeme τ da uređaj reagira.

     → vrijeme reakcije uvijek uzrokuje kašnjenje mjerenja za isti iznos τ

  c) Kut očitavanja daje krivi ravnotežni položaj

     → kut očitavanja uzrokuje pogrešno mjerenje vremena za isti iznos t

Sistematske pogreške (II)

***Netočnosti mjerenja* uzrokovane mjernim uređajem ili tehnikom**

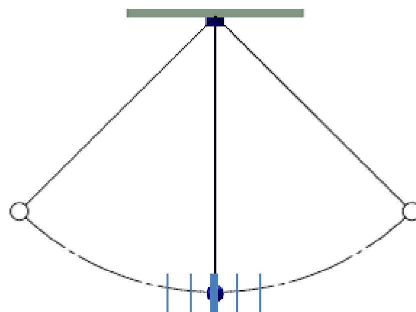
→ npr. pogrešno kalibrirana ili pomaknuta skala

- Prilikom ponavljanja mjerenja javljaju se u istom smjeru i iznosu (reproducibilnost!)
- Nisu predmetom statističke analize
- Mogu se ukloniti ili smanjiti na nekoliko načina:
  - poboljšanjem aparature ili tehnike – npr. preciznija kalibracija
  - planiranjem mjerenja – mjerenje se u nekim situacijama može organizirati tako da se sistematske pogreške ponište.
    Primjer: kada mjerimo vrijeme štopericom svaki puta kad pritisnemo gumb mjerenja kasni za vrijeme τ. U slučaju mjerenja perioda njihala isto vrijeme kašnjenja je prisutno pri pokretanju i zaustavljanju štoperice, pa se ove dvije pogreške poništavaju.

Slučajne pogreške (I)

<u>Primjer</u>:

- Mjerimo period titranja njihala pomoću štoperice
  → u tom slučaju **mjerni uređaj** čine čovjek + štoperica



- Izvršimo 10 mjerenja i dobivamo vrijednosti koje se razlikuju
- Što može dovesti do razlike u rezultatima mjerenja:
  a) Brzine reakcije kod uključivanja/zaustavljanja štoperice
  b) Preciznost štoperice (zaokruživanje decimala)
  c) Nejednako očitavanje ravnotežnog položaja njihala
  → nesavršenost mjernog uređaja
  d) Promjene ravnotežnog položaja njihala u vremenu, npr. zbog drmanja postolja na kojem se nalazi ili strujanja zraka
  → utjecaj okoline

Mjerenje	T [s]
1.	12.5
2.	11.9
3.	13.0
4.	13.1
5.	12.6
6.	12.2
7.	12.7
8.	11.6
9.	13.4
10.	12.8

Slučajne pogreške (II)

- Svojstva slučajnih pogrešaka:
  - Ponavljanjem mjerenja dobivaju se različiti rezultati - slučajne pogreške različite po iznosu i smjeru
  - Mogu se statistički obraditi (čime ćemo se mi baviti)
- Uzrok pogreške su nestalni uvjeti mjerenja:
  - Preciznost mjernog uređaja
  - Promjena okoline
- Smanjenje slučajnih pogrešaka:
  - Usavršavanje mjernog uređaja ili tehnike (npr. automatizacija mjerenja u našem primjeru)
  - Izolacija od okoline
  - Ponavljanjem mjerenja i statističkom obradom može se preciznije odrediti prava vrijednost fizikalne veličine

Slučajnu pogrešku možemo definirati kao *neodređenost rezultata* zbog konačne preciznosti uređaja i fluktuacija u uvjetima mjerenja.

Neovisna mjerenja

- Izvodimo niz mjerenja neke fizikalne veličine ***X*** i dobivamo rezultate $x_1, x_2, x_3, \ldots, x_n$, koji se međusobno razlikuju zbog prisustva slučajnih pogrešaka
- Da bi odredili najvjerojatniju vrijednost mjerene veličine i pogrešku mjerenja definiramo sljedeće pojmove:
  - Srednja vrijednost
  - Srednja kvadratna pogreška pojedinog mjerenja
  - Srednja kvadratna pogreška aritmetičke sredine
  - Relativna nepouzdanost
  - Maksimalna apsolutna pogreška

Srednja vrijednost

- izračunava se kao aritmetička sredina izmjerenih vrijednosti
- za *n* mjerenja aritmetička sredina je:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Uzimamo da je upravo $\bar{x}$ najvjerojatnija prava vrijednost ***X*** mjerene fizikalne veličine → $\bar{x}$ govori o očekivanoj vrijednosti mjerene veličine

Srednja kvadratna pogreška mjerenja (standardna devijacija)

- kvadratna odstupanja izmjerenih vrijednosti od srednje vrijednosti

$$m = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

- za dovoljno velik n (~10 mjerenja), m poprima ustaljenu vrijednost
- govori o pouzdanosti pojedinog mjerenja
- iskazuje rasipanje rezultata kao posljedicu preciznosti uređaja → *mjera preciznosti uređaja*

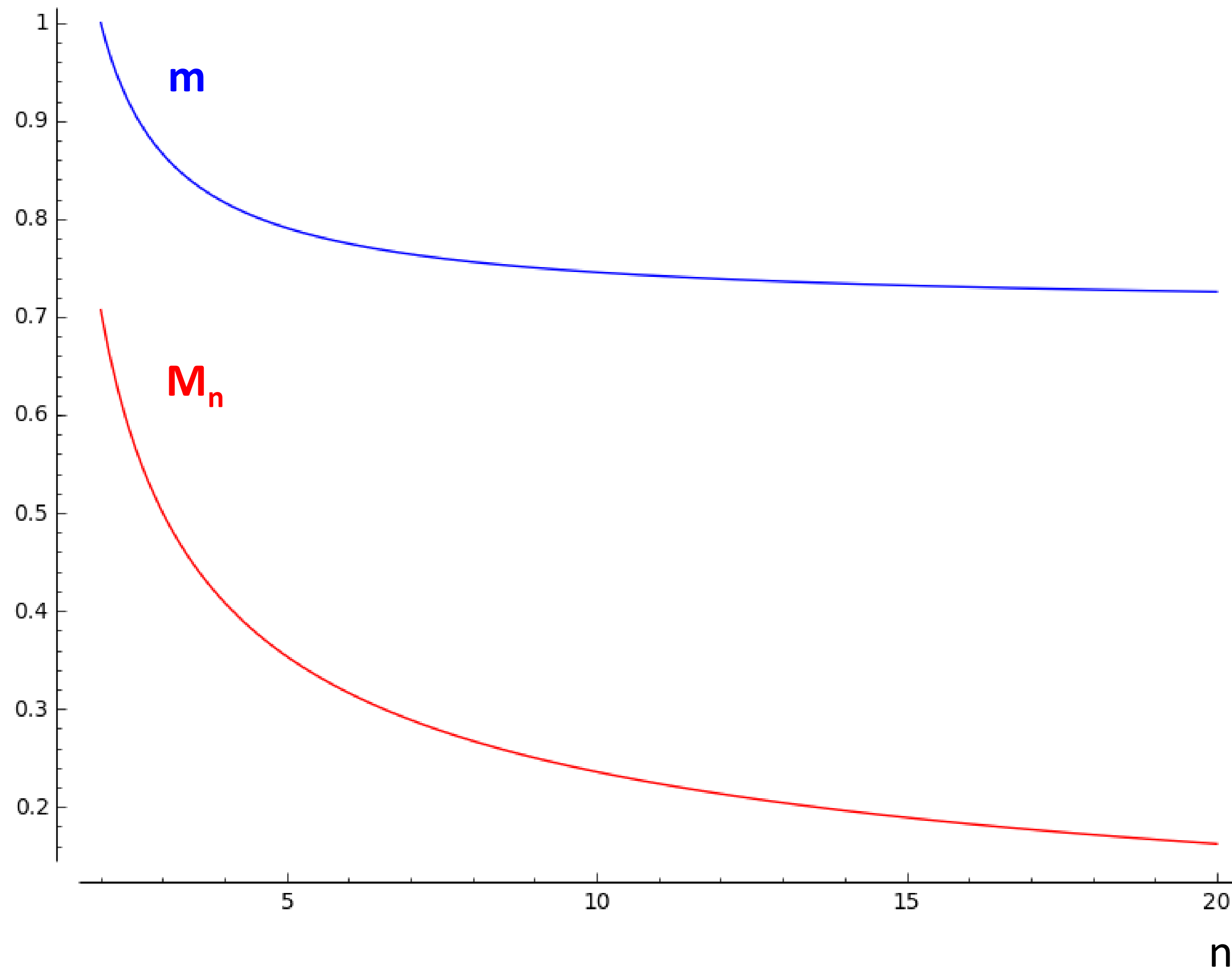
Srednja kvadratna pogreška aritmetičke sredine (nepouzdanost)

- Ako izvedemo veći broj mjerenja očekujemo da će fizikalna veličina biti preciznije određena
- Mjera *preciznosti rezultata* je srednja kvadratna pogreška aritmetičke sredine:

$$M_n = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n(n-1)}}$$

- M_n se smanjuje sa brojem mjerenja proporcionalno ~ $1/\sqrt{n}$ → povećava se preciznost rezultata
- Govori o ***pouzdanosti rezultata***

Pogreške u odnosu na broj mjerenja



Maksimalna apsolutna pogreška

- najveće odstupanje pojedinačnog mjerenja od aritmetičke sredine:

$$\Delta x = |\bar{x} - x_i|_{max}$$

- Ponekad zbog prirode eksperimenta ne možemo izračunati M_n:
  - ako imamo samo jedno mjerenje
  - ako sva mjerenja daju isti rezultat pa je M_n=0
- Tada procjenjujemo maksimalnu pogrešku Δx

Relativna nepouzdanost

- ako je poznat M_n:

$$R = \frac{M_n}{\bar{x}} \cdot 100\%$$

- ako nije poznat M_n :

$$R = \frac{\Delta x}{\bar{x}} \cdot 100\%$$

Rezultat

- pišemo u obliku:

$$x = \bar{x} \pm M_n$$

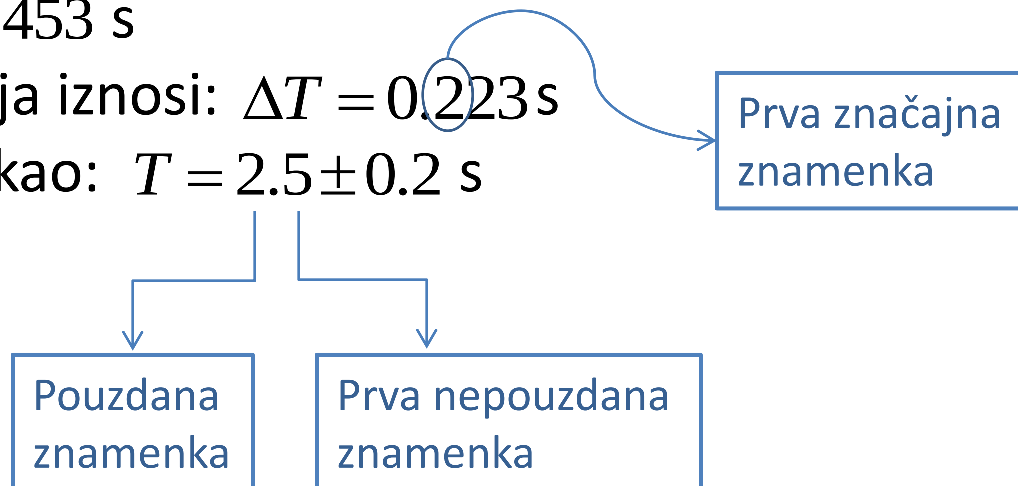
- ako nije poznat M_n:

$$x = \bar{x} \pm \Delta x$$

Zaokruživanje rezultata

Primjer:

– pri ponavljanju mjerenja perioda njihala dobili smo srednju vrijednost: $T = 2.453$ s
– pogreška mjerenja iznosi: $\Delta T = 0.223$ s
– Rezultat pišemo kao: $T = 2.5 \pm 0.2$ s



Pogreška se u pravilu zaokružuje na prvu znamenku različitu od nule

Rezulat se zaokružuje na prvu nepouzdanu znamenku – onu koja je na istom decimalnom mjestu kao i zaokružena pogreška

Primjer

- Mjerenje perioda ***T*** matematičkog njihala
- Na temelju izmjerenih vrijednosti dobivamo:

$$\bar{T} = 12{,}6\ s$$

$$m = 0{,}6\ s$$

$$M_{10} = 0{,}2\ s$$

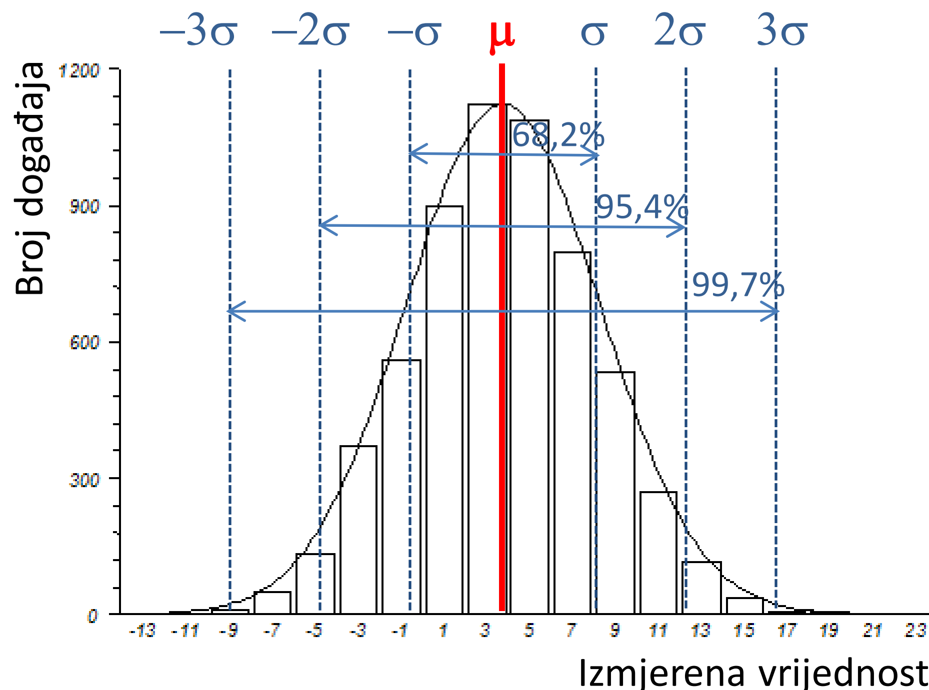
- Pa pišemo rezultat mjerenja:

$$T = 12{,}6 \pm 0{,}2\ s$$

Mjerenje	T [s]
1.	12,5
2.	11,9
3.	13,0
4.	13,1
5.	12,6
6.	12,2
7.	12,7
8.	11,6
9.	13,4
10.	12,8

Gaussova raspodjela

- Ako ponavljamo mjerenje veličine *x*, dobit ćemo niz rezultata $x_1, x_2, x_3, \ldots, x_n$, koji se međusobno razlikuju zbog prisustva slučajnih pogrešaka
- Može se pokazati da vrijednosti dobivene ponavljanjem mjerenja slijede ***Gaussovu*** raspodjelu. (Pri tome zanemarujemo sistematske greške.)



Rezultat mjerenja prikazan kao:

$$x = \bar{x} \pm M_x$$

srednja vrijednost — nepouzdanost

može se opisati pomoću Gaussove raspodjele: $\sim e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

gdje pretpostavljamo: $\bar{x} = \mu$

$$M_x = \sigma / \sqrt{n}$$

Ovisna mjerenja

- Tražena veličina *F* je funkcija neposredno izmjerenih veličina x_i, $F = f(x_1, \dots, x_i, \dots x_n)$
- Najvjerojatnija vrijednost je srednja vrijednost:

$$\bar{F} = f(\bar{x}_1, \dots, \bar{x}_i, \dots \bar{x}_n)$$

- Ako su veličine x_i međusobno neovisne onda je srednja kvadratna pogreška veličine *F*:

$$M_F = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}\bigg|_{\bar{x}} M_i \right)^2}$$

- Rezultat pišemo kao: $F = \bar{F} \pm M_F$

Ovisna mjerenja

Primjer: želimo odrediti ubrzanje sile teže mjerenjem perioda titranja (*T*) i duljine niti (*l*) matematičkog njihala

- Ubrzanje sile teže je dano relacijom: $g = 4\pi^2 \frac{l}{T^2}$
- Uzmimo da su neovisnim mjerenjima dobiveni rezultati:
  - ***l*** = 0.850 ± 0.002 m
  - ***T*** = 1.849 ± 0.003 s
- Srednje ubrzanje sile teže je: $\bar{g} = 4\pi^2 \frac{\bar{l}}{\bar{T}^2}$
- Srednja kvadratna pogreška je:

$$M_g = \sqrt{\left(\frac{\partial g}{\partial l} M_l\right)^2 + \left(\frac{\partial g}{\partial T} M_T\right)^2} \qquad M_g = \sqrt{\left(\frac{4\pi^2}{\bar{T}^2} M_l\right)^2 + \left(2\frac{4\pi^2 \bar{l}}{\bar{T}^3} M_T\right)^2}$$

- Uvrštavanjem dobivamo: g=(9.82 ±0.04) ms^{-2}

Opća srednja vrijednost

- Izvedeno je ***n*** nizova mjerenja **iste** fizikalne veličine te je za svaki niz dobivena srednja vrijednost i kvadratna pogreška:

$$x_i = \bar{x}_i \pm M_i, \quad i = 1 \ldots n$$

→ primjer fizikalna veličina je određena različitim eksperimentalnim metodama

- **Konzistentna mjerenja** – ako su razlike za svaki par mjerenja $|\bar{x}_i - \bar{x}_j|$ usporedive s bilo kojim M_k
- **Nekonzistentna mjerenja** - ako su razlike $|\bar{x}_i - \bar{x}_j| >> M_k$ tada zanemarujemo nepouzdanosti M_k, a veličine $\bar{x}_k$ smatramo nezavisnama te ih tako i analiziramo

Opća srednja vrijednost

- Za **konzistentna mjerenja** definiramo opću aritmetičku sredinu:

$$\bar{x} = \left(\sum_{i=1}^{n} \frac{\bar{x}_i}{M_i^2}\right) M^2$$

- Gdje je *M* nepouzdanost:

$$M = \frac{1}{\sqrt{\sum_{i=1}^{n} M_i^{-2}}}$$

- Rezultat pišemo u obliku:

$$x = \bar{x} \pm M$$

- Poseban slučaj **konzistentnih mjerenja** je kada je jedna pouzdanost M_k znatno manja od svih ostalih. Tada vrijedi $M \approx M_k$

Povećanje preciznosti rezultata
ili kako smanjiti pogreške?

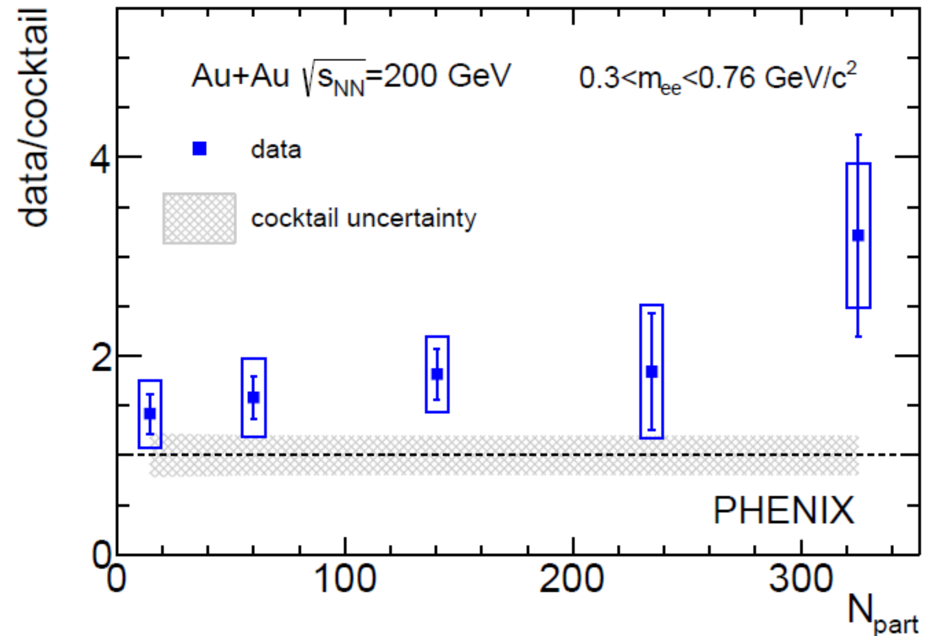
- Moguće sistematske pogreške treba reducirati pri planiranju eksperimenta – treba razviti i napraviti mjerni uređaj ili tehniku koja će omogućiti relativno male sistematke pogreške
  → bolji uređaj u pravilu znači skuplji uređaj
- Slučajne pogreške u pravilu se mogu smanjiti ponavljanjem mjerenja
  → dulje mjerenje znači skuplje mjerenje
- Koliko mjerenja treba napraviti? Toliko da slučajna (statistička) pogreška bude manja ili podjenaka sistematskoj
  → najčešče se pokazuje da je to financijski i vremenski najefikasnije rješenje

Statističke i sistematske pogreške

- Primjer mjerenja

→ Svaka izmjerena točka ima pripadajuću statističku pogrešku (vertikalne crte) i sistematsku pogrešku (pravokutnici)

→ Obje pogreške su usporedive



A. Adare *et al.* (PHENIX Collaboration), Phys. Rev. C 93 014904

Statistika i osnovna mjerenja

Grafički prikaz rezultata

M. Makek

2016/2017

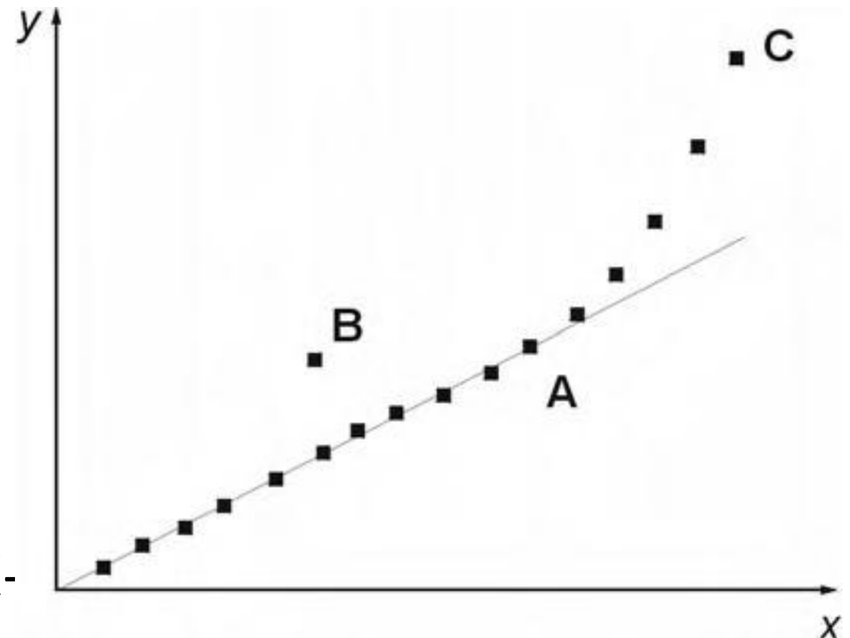
GRAFIČKI PRIKAZ REZULTATA

X-Y graf

Histogram

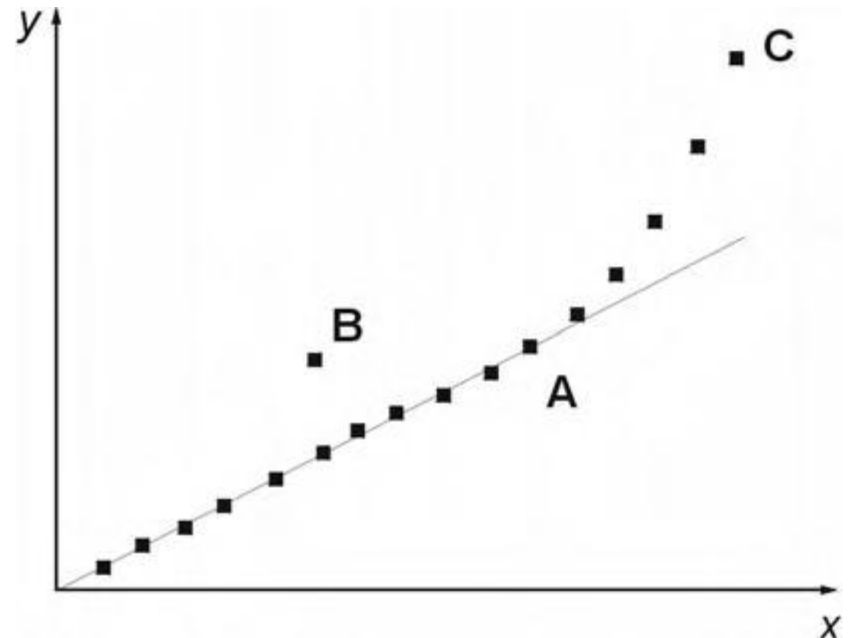
X-Y graf

- Pretpostavimo da smo mjerenjem fizikalnih veličina dobili niz točaka (x_i, y_i)
- Uobičajeno je da ***x*** ona veličina koju preciznije (neovisno) mjerimo te je unosimo na apscisu
- Izmjerene točke crtamo na X-Y graf (eng. scatter plot)



X-Y graf

- Što uočavamo na grafu:
  a) Linearnost od ishodišta do točke A
  b) Nelinearnost od točke A do točke C
  c) Rasipanje točaka oko zamišljenog pravca u linearnom području → ocjena slučajnih pogrešaka
  d) Veliko odstupanje točke B – moguće da se radi o gruboj pogreški, eventulano se može zanemariti.
     Ali: velika odstupanja na rubovima grafa poput točke C se ne smiju zanemariti jer mogu ukazivati na novu pojavu



X-Y graf

- Prednosti grafičkog nad isključivo numeričkim prikazom:
  - zorno se vidi kako jedna veličina ovisi o drugoj – mogu se uočavati zakonitosti, kao npr. linearna ili nelinearna ovisnost na prethodnom primjeru
  - vidi se koliko pojedino mjerenje odstupa od trenda - lako se uočavaju grube pogreške (poput točke B na prethodnom primjeru)
- Pri tome treba paziti na:
  - Pogodno mjerilo na apscisi i ordinati – graf treba biti popunjen
  - Odabir varijabli za crtanje – ako je moguće linearizirati (o tome kasnije)
  - Označiti fizikalne veličine i mjerne jedinice na svakoj osi i pri tome koristiti dovoljno velika slova
  - Kod opisa grafa treba jasno naznačiti što je mjereno i što je prikazano

X-Y graf

- Interpolacija
  - Može se procijeniti zavisna varijabla *y* u točki koja nije izmjerena, a nalazi se između dvije mjerene točke
  - U pravilu daje ispravne vrijednosti, ali treba uzeti u obzir razlučivost
- Ektrapolacija
  - Može se procijeniti zavisna varijabla *y* u točkama koje se protežu izvan mjernog područja
  - U pravilu se izbjegava jer u nepoznatom području može doći do odstupanja od trenda – npr. odstupanje od pravca između točaka A i C u prethodnom primjeru

Linearizacija X-Y grafa

- Nakon unošenja točaka u graf uočavamo da ovisnost *y(x)* nije linearna, tj. točke ne leže na pravcu

- Ako očekujemo nelinearnu ovisnost onda pokušamo pravilnim odabirom varijabli na x i y osi nacrtati linearan graf.

  Primjer: mjerimo ovisnost prijeđenog puta o vremenu pri jednoliko ubrzanom gibanju

  → očekujemo ovisnost: $s(t) = 1/2at^2$

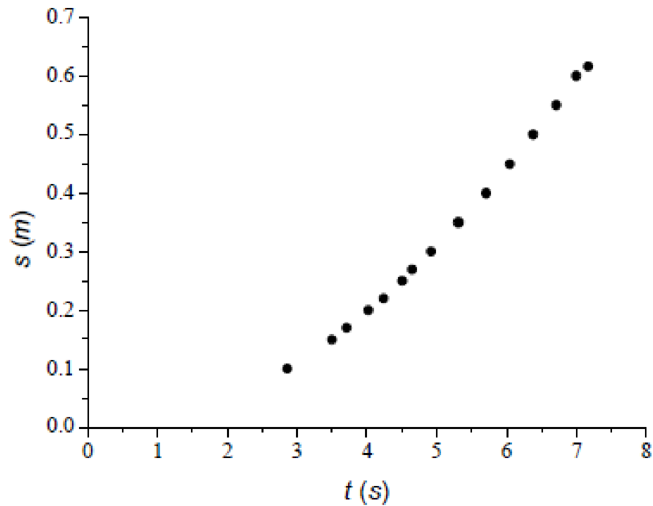
  → supstitucija: $u=t^2$

  → Crtamo graf: ***s(u) - u*** za koji očekujemo da će ovisnost biti linearna

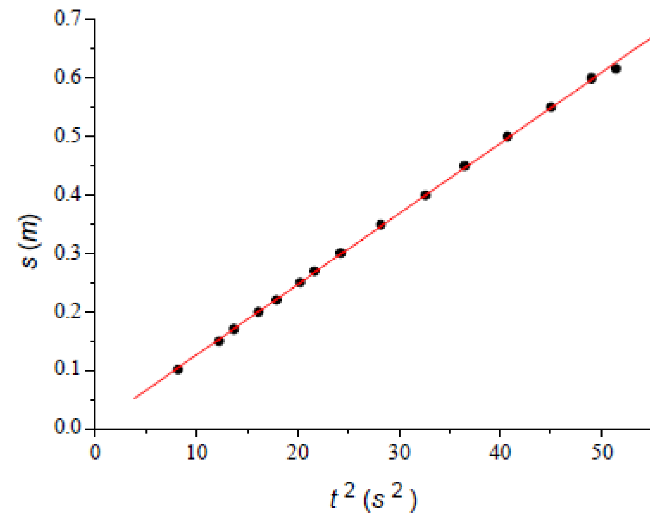
Linearizacija X-Y grafa

Primjer 1:

Kvadratna ovisnost ***s(t) = 1/2at²*** → *linearna ovisnost* ***s(t²)=s(u)=1/2a u***



$u = t^2$



Linearizacija X-Y grafa

- Ako uočimo nelinearnu ovisnost točaka, a točna funkcijska ovisnost *y(x)* nam nije poznata možemo iskoristiti **pravilo logaritmiranja:**
  - za funkcije tipa $y = ax^b$ logaritmiranjem dobivamo: $\ln y = \ln a + b \ln x$
    → crtamo graf ln *y* – ln *x* u kojem će ***b*** odgovarati nagibu pravca a **ln *a*** odsječku na osi y
  - Za funkcije tipa $y = ae^{xb}$ logaritmiranjem dobivamo: $\ln y = \ln a + bx$
    → crtamo graf ln *y* – *x* u kojem će ***b*** odgovarati nagibu pravca a **ln *a*** odsječku na osi y

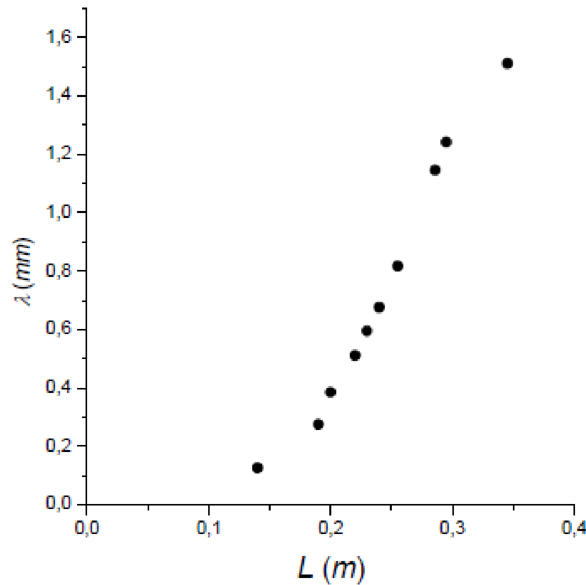
Linearizacija X-Y grafa

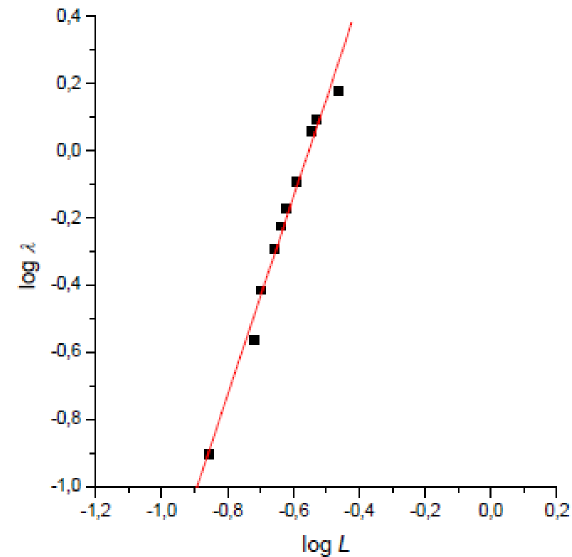
<u>Primjer 2:</u>

Nepoznata ovisnost tipa $\boldsymbol{\lambda(L)=AL^{\alpha}}$ → *logaritmiranje*

→ *linearna ovisnost* $\boldsymbol{\log \lambda = \log A + \alpha \log L}$

$x = \log L \qquad y = \log \lambda$





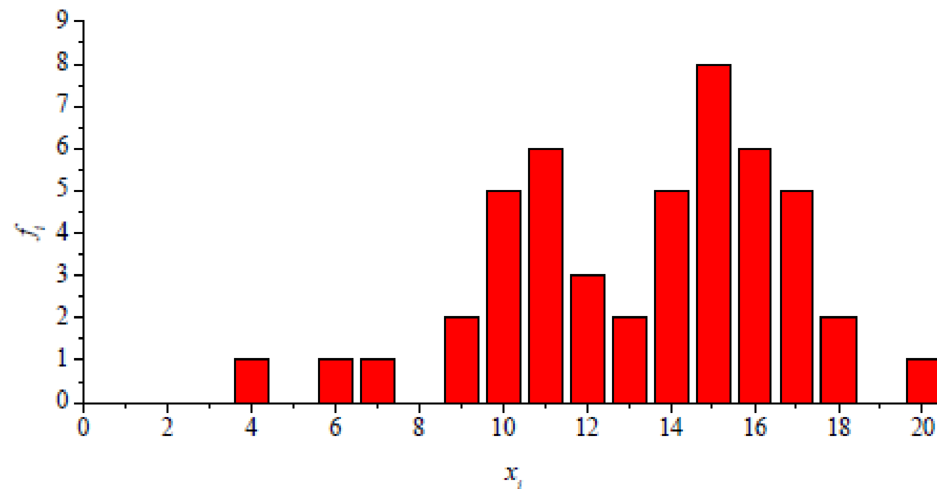
Histogram

- Graf koji prikazuje raspodjelu frekvencija diskretne ili kontinuirane varijable
  - Diskretne varijable – skup vrijednosti prebrojiv
  - kontinuirane varijable – skup vrijednosti kontinuiran
- Frekvencija f_i neke određene vrijednosti x_i varijable X jest broj pojavljivanja te vrijednosti u promatranom skupu podataka
- Relativna frekvencija je definirana kao: $f_{ri} = f_i/N$, gdje je N ukupan broj podataka

x_i	f_i
1	0
2	0
3	0
4	1
5	0
6	1
7	1
8	0
9	2
10	5
11	6
12	3
13	2
14	5
15	8
16	6
17	5
18	2
19	0
20	1

Crtanje histograma

- Za diskretnu varijablu:
  - Odrediti frekvencije ili relativne frekvencije
  - Nacrtati moguće vrijednosti fizikalne veličine na x osi
  - Nacrtati odgovarajuće frekvencije na y-osi
- Primjer:

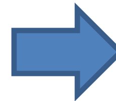


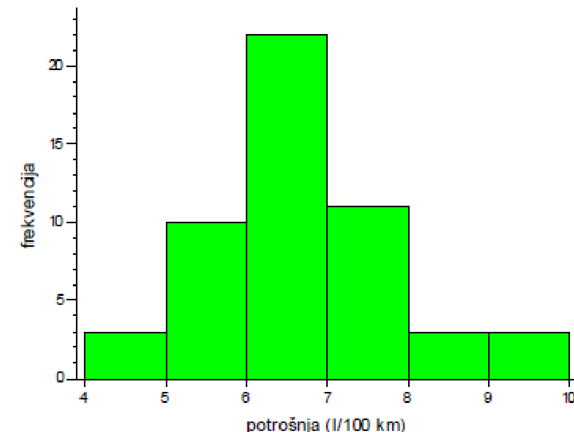
Crtanje histograma

- Za kontinuiranu varijablu:
  - Podijeliti interval mjerenja veličine X na konačan broj razreda (mogu biti ekvidistantni ili neekvidistantni)
  - Grupirati podatke u razrede i odrediti (relativne) frekvencije
  - Nacrtati razrede na x osi
  - Nacrtati odgovarajuće frekvencije na y-osi
- Primjer:

potrošnja goriva

6,25	5,93	7,8	4,95	9,2	8,57
6,82	7,43	5,78	5,46	6,54	7,02
6,78	4,75	5,32	7,11	5,66	5,99
6,87	8,35	7,66	7,23	6,58	6,92
6,32	7,08	5,98	6,25	5,45	6,72
6,38	6,9	9,87	6,23	6,52	6,43
6,12	5,81	6,37	7,23	7,46	8,06
6,09	5,82	4,99	6,32	6,51	6,49
9,49	6,39				





Statistika i osnovna mjerenja

Metoda najmanjih kvadrata

M. Makek

2016/2017

METODA NAJMANJIH KVADRATA

Linearna regresija

Linearizacija ovisnosti

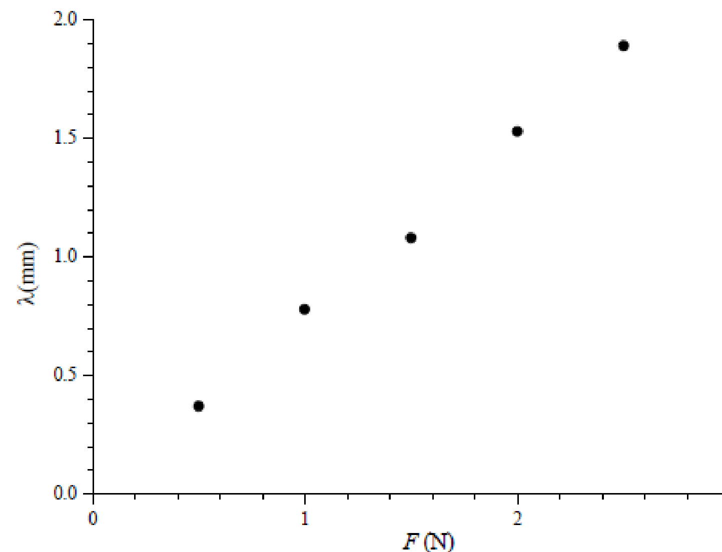
Nelinearna regresija

Linearna regresija
(veza zavisne i nezavisne varijable linearna)

- Primjer:
  - Mjerimo ovisnost produljenja opruge (λ) o primjenjenoj sili (F)
  - Izmjereni su podaci te je nacrtan graf:

F(N)	λ(mm)
0,5	0,37
1	0,78
1,5	1,08
2	1,53
2,5	1,89





- Čini se da točke imaju linearnu ovisnost. Kako odrediti pravac koji "najbolje" slijedi točke?

Metoda najmanjih kvadrata

kod linearne regresije

- Od nekog pravca ***y=ax+b*** točka (x_i, y_i) odstupa za ε_i odnosno vrijedi možemo pisati: $y_i = ax_i + b + \varepsilon_i$
- Princip najmanjih kvadrata:

Od svih pravaca regresije najvjerojatniji je onaj za koji je suma kvadrata odstupanja minimalna

$$S(a,b) \stackrel{\text{def}}{=} \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} [y_i - (ax_i + b)]^2 = min$$

- odnosno $\dfrac{\partial S(a,b)}{\partial a} = 0 \quad \dfrac{\partial S(a,b)}{\partial b} = 0$

Metoda najmanjih kvadrata

kod linearne regresije

$$\frac{\partial S(a,b)}{\partial a} = 0$$

$$\frac{\partial S(a,b)}{\partial a} = \sum_i 2[y_i - (ax_i + b)](-x_i)$$

$$= \sum_i 2\left[-x_i y_i + \left(ax_i^2 + bx_i\right)\right]$$

$$= 2\sum_i -x_i y_i + 2a\sum_i x_i^2 + 2b\sum_i x_i$$

$$a\sum_i x_i^2 + b\sum_i x_i - \sum_i x_i y_i = 0 \qquad (1)$$

Metoda najmanjih kvadrata

kod linearne regresije

$$\frac{\partial S(a,b)}{\partial b} = 0$$

$$\frac{\partial S(a,b)}{\partial b} = \sum_i 2[y_i - (ax_i + b)]$$

$$= 2\sum_i y_i - 2a\sum_i x_i - 2b\sum_i 1$$

$$\sum_i y_i - a\sum_i x_i - bn = 0$$

$$b = \frac{\sum_i y_i - a\sum_i x_i}{n} \qquad (2)$$

Metoda najmanjih kvadrata

kod linearne regresije

- Uvrštavanjem izraza (2) u izraz (1) i sređivanjem dobivamo izraze za koeficijente pravca:

$$a = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

$$b = \frac{1}{n}\left(\sum y_i - a \sum x_i\right)$$

Metoda najmanjih kvadrata

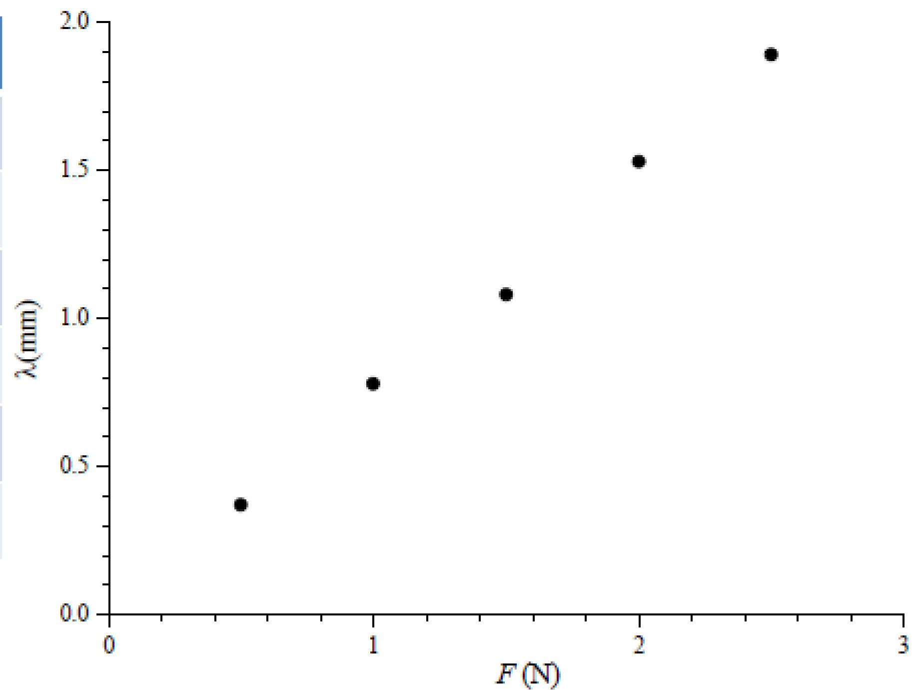
- Nepouzdanosti koeficijenata slijede iz izraza za *a* i *b* te izraza za propagaciju pogreške kod zavisnih mjerenja:

$$M_a = \sqrt{\frac{1}{(n-2)}\left[\frac{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} - a^2\right]}$$

$$M_b = M_a\sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

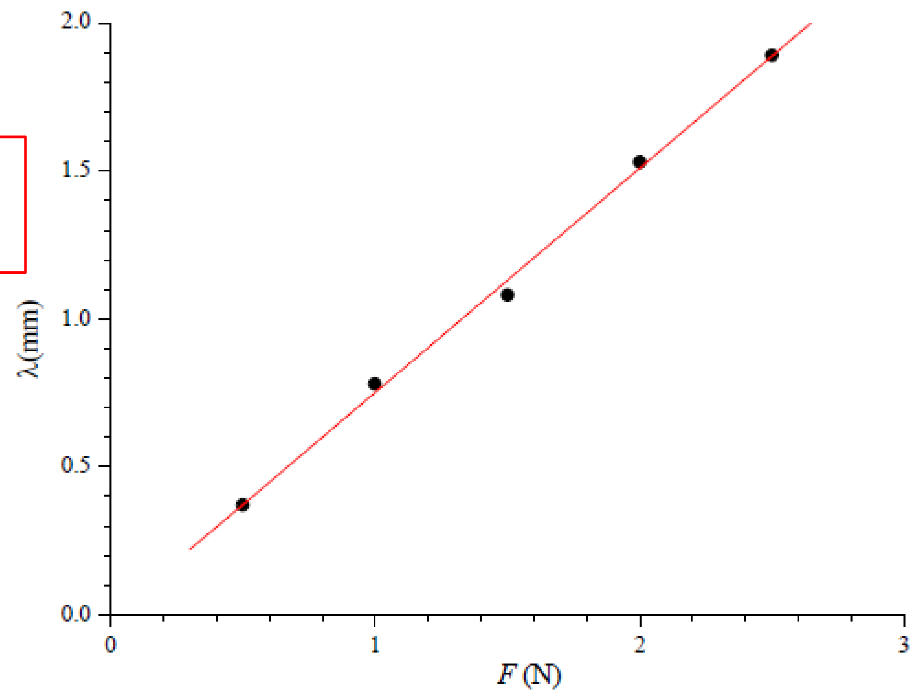
Metoda najmanjih kvadrata

i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	0.5	0.37	0.25	0.137	0.185
2	1.0	0.78	1.0	0.608	0.78
3	1.5	1.08	2.25	1.166	1.62
4	2.0	1.53	4.0	2.341	3.06
5	2.5	1.89	6.25	3.572	4.72
Σ	**7.5**	**5.65**	**13.75**	**7.824**	**10.365**



Metoda najmanjih kvadrata

a = 0.756 mm/N M_a = 0.038 mm/N
b =- 0.004 mm M_b = 0.063 mm

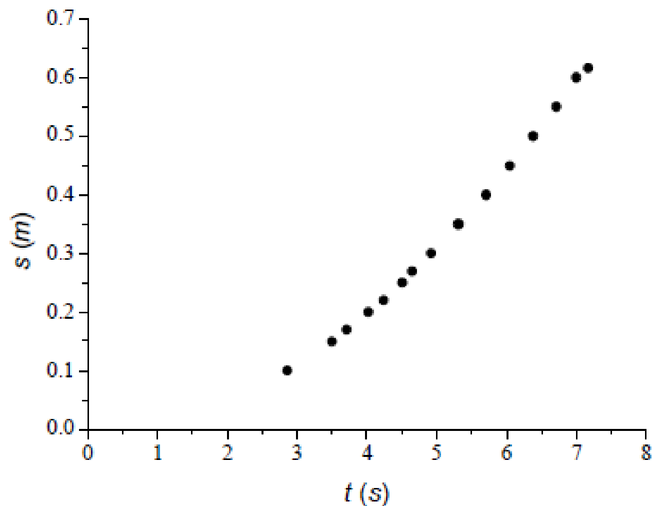


Linearizacija

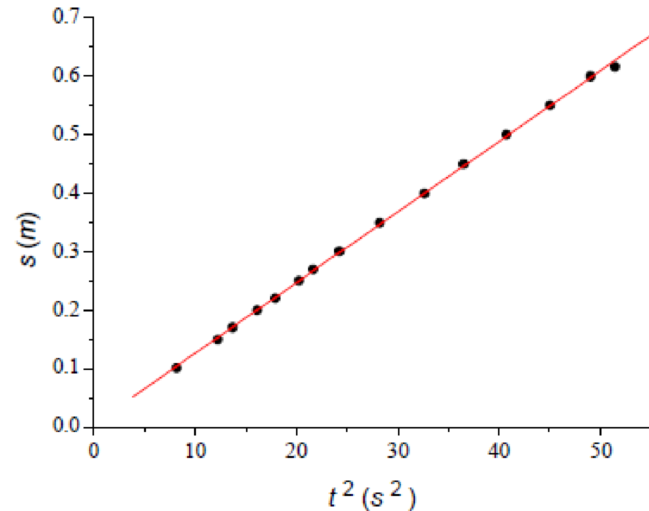
<u>Primjer 1:</u>

Kvadratna ovisnost $\boldsymbol{s(t) = 1/2at^2}$ → *linearna ovisnost* $\boldsymbol{s(t^2)=s(u)=1/2a\,u}$

→ dalje se računa s(u) kao da se radi o linearnoj regresiji



$u = t^2$



<u>Primjer 2</u>: linearizacija se vrši logaritmiranjem relacije → dalje račun za linearnu regresiju.

Nelinearna regresija

- u slučaju kada zavisna varijabla nelinearno ovisi o nezavisnoj
- primjer – tjerani prigušeni harmonički oscilator:

$$y(\omega) = \frac{A}{\sqrt{(\omega_0^2 - \omega^2)^2 + \left(\frac{\omega}{\tau}\right)^2}}$$

- potrebno je odrediti parametre *A,* ω_0 *i* τ, za koje je suma kvadrata odstupanja minimalna:

$$S(A, \omega_0, \tau) = \sum_i \left(y_i - \frac{A}{\sqrt{(\omega_0^2 - \omega^2)^2 + \left(\frac{\omega}{\tau}\right)^2}} \right)^2$$

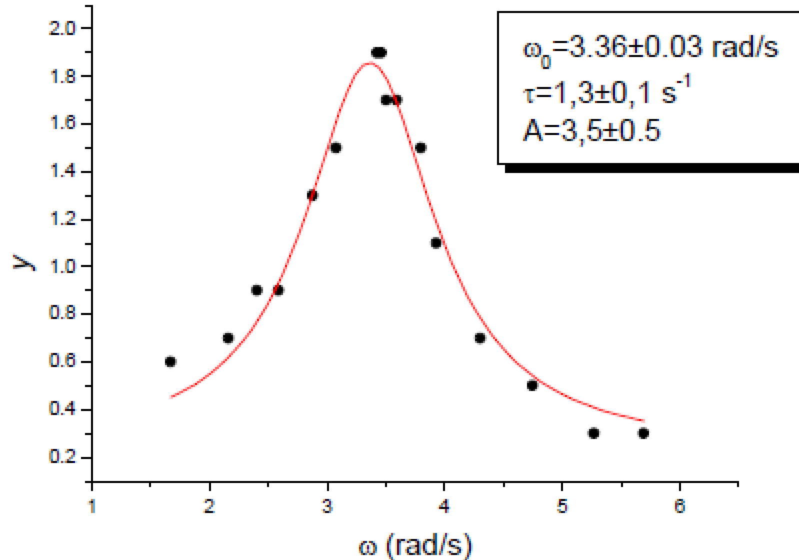
- Iz uvjeta minimizacije:

$$\frac{\partial S(A, \omega_0, \tau)}{\partial A} = 0 \qquad \frac{\partial S(A, \omega_0, \tau)}{\partial \omega_0} = 0 \qquad \frac{\partial S(A, \omega_0, \tau)}{\partial \tau} = 0$$

- Slijede koeficijenti *A,* ω_0 *i* τ

Nelinearna regresija

- primjer



$$y = \frac{A}{\sqrt{\left(\omega_0^2 - \omega^2\right)^2 + \left(\omega / \tau\right)^2}}$$

Primjer: matematičko njihalo - mjerenje ovisnosti perioda titranja o duljini niti

https://phet.colorado.edu/sims/pendulum-lab/pendulum-lab_en.html

Zadatak:

a) Izmjerite period (T) za 5 različitih duljina (L) i napravite tablicu L-T
b) Linearizirajte ovisnost i metodom najmanjih kvadrata odredite ubrzanje sile teže
c) Logaritmirajte ovisnost i odredite α u izrazu $T \sim L^{\alpha}$

Statistika i osnovna mjerenja

Osnove kombinatorike

M. Makek

2016/2017

OSNOVNI POJMOVI KOMBINATORIKE

Teorem o uzastopnom prebrojavanju

Permutacije

Varijacije

Kombinacije

Teorem o uzastopnom prebrojavanju

Kombinatorijski problem:

- Želimo popuniti prazna mjesta 1 i 2
- Imamo n_1 elemenata kojima možemo popuniti mjesto 1 ($A_1,\dots, A_{n1}$)
- Imamo n_2 elemenata kojima možemo popuniti mjesto 2 ($B_1,\dots, B_{n2}$)
- Na koliko se različitih načina mogu popuniti prazna mjesta 1 i 2?

	A_1	A_2	…	A_{n1}
B_1	A_1B_1	A_2B_1	…	$A_{n1}B_1$
B_2	A_1B_2	A_2B_2	…	$A_{n1}B_2$
…	…	…	…	…
B_{n2}	A_1B_{n2}	A_2B_{n2}	…	$A_{n1}B_{n2}$

→ Može se kombinirati svaki A sa svakim B, stoga imamo $\mathbf{n_1n_2}$ načina

Teorem o uzastopnom prebrojavanju

- Pretpostavimo da imamo k praznih mjesta
- Imamo n_1 elemenata kojima možemo popuniti mjesto 1 ($A_1,\ldots, A_{n1}$)
- Imamo n_2 elemenata kojima možemo popuniti mjesto 2 ($B_1,\ldots, B_{n2}$)
- Imamo n_k elemenata kojima možemo popuniti mjesto k ($Y_1,\ldots, Y_{nk}$)

TM: k mjesta možemo popuniti na $n_1 n_2 \ldots n_k$ načina

→ Dokaz teorema je indukcijom iz prethodnog primjera

<u>Primjer</u>:

- 2 predjela
- 3 glavna jela
- 3 deserta

} Možemo složiti 18 različitih jelovnika

Permutacije bez ponavljanja

- Pretpostavimo da imamo **n različitih elemenata A_1,..., A_n** koje možemo poredati u niz na više načina → svaki takav niz zovemo **permutacija**
- Koliko ima permutacija n elemenata?
  - Ekvivalentno pitanje je: na koliko se načina može popuniti n mjesta s n elemenata?

- ❑ 1. mjesto možemo popuniti na n načina → preostaje n-1 elemenata
- ❑ 2. mjesto možemo popuniti na n-1 načina → preostaje n-2 elemenata
- ❑ k. mjesto možemo popuniti na n-(k-1) način
- ❑ n. mjesto možemo popuniti na n-(n-1) = 1 način

Prema teoremu o uzastopnom prebrojavanju n mjesta mogu se popuniti na:

$$P^{(n)} = n(n-1)(n-2)\ldots 3 \cdot 2 \cdot 1 \overset{\text{def}}{=} n!$$ (čitamo: n faktorijela)

po dogovoru je: 0!=1

Permutacije bez ponavljanja

Primjer: koliko riječi možemo složiti iz znakova **A, P, N**?

- Prvo mjesto možemo odabrati na 3 načina:
  - A _ _
  - P _ _
  - N _ _
- Drugo mjesto možemo odabrati na 2 načina:
  - A P _ , A N _
  - P A _ , P N _
  - N A _ , N P _
- Treće mjesto na jedan način:
  - A P N , A N P
  - P A N , P N A
  - N A P , N P A
- Ukupno 3! = 6 načina

Permutacije s ponavljanjem

- Pretpostavimo da imamo **n elemenata $A_1, \ldots, A_n$** od kojih je $r_1, r_2, \ldots, r_k$ identičnih. Koliko ima permutacija takvih n elemenata?
- Primjer: koliko ima permutacija niza AABBB?

AABBB	BAABB	BBAAB	BBBAA
ABABB	BABAB	BBABA	
ABBAB	BABBA		
ABBBA			

Za svaki od ovih nizova postoje:
- 2! identična koji se dobiju permutacijama A
- 3! identična koji se dobiju permutacijama B

- $n=5$, $r_1=2$, $r_2=3$

  → broj različitih permutacija: n! / (r1!r2!) = 5! / (2!3!) = 10

Permutacije s ponavljanjem

- Ako imamo r_1 identičnih elemenata znači da se njihovom zamjenom ne dobivaju nove permutacije -- > dobivamo grupe od po $r_1!$ identičnih permutacija. Broj takvih grupa je $n!/r_1!$
- Od preostalih $n-r_1$ elemenata imamo r_2 identičnih. Taj skup permutacija dijeli se na grupe od po $r_2!$ identičnih permutacija. Broj permutacija je $n!/ r_1! r_2!$
- Općenito je onda broj permutacija $r_1, \ldots, r_k$ jednakih elemenata:

$$P^{(n)}{}_{r_1,r_2,\ldots,r_k} = \frac{n!}{r_1!r_2!\ldots r_k!}$$

- Ovo je poopćenje relacije za broj permutacija

  → Ako su $r_1, r_2, \ldots r_k = 1$ (ili 0) onda se ova relacija svodi na slučaj permutacija bez ponavljanja

Varijacije bez ponavljanja

- Pretpostavimo da imamo **n različitih elemenata $A_1,\dots, A_n$** koje želimo **poredati** u niz od **r** članova, ne dopuštajući ponavljanje jednog te istog elementa A_k → svaki takav niz zovemo **varijacija n-tog reda i r-tog razreda**
- Koliko ima takvih varijacija?
- Zamislimo da treba popuniti niz od **r** članova s **n** različitih elemenata (pri tome je uvijek r≤n):
  - 1. mjesto možemo popuniti na n načina
  - 2. mjesto možemo popuniti na n-1 način
  - r. mjesto možemo popuniti na n-(r-1) način
- Slijedi da je broj varijacija n-tog reda i r-tog razreda:

$$V_r^{(n)} = n(n-1)\dots(n-r+1) \quad / \quad \frac{(n-r)!}{(n-r)!}$$

$$V_r^{(n)} = \frac{n!}{(n-r)!}$$

Varijacije bez ponavljanja

- Primjer:
  - Imamo na raspolaganju 27 slova
  - Želimo odabrati oznake od 2 slova za registarske tablice
  - Koliko oznaka možemo napraviti?

  → Obzirom da ovdje razlikujemo poredak slova i ne dopuštamo ponavljanje slova, radi se o varijacijama 27. reda i 2. razreda

  →V = 27!/25!=702

Varijacije s ponavljanjem

- Pretpostavimo da imamo **n različitih elemenata $A_1,\dots, A_n$** koje želimo **poredati** u niz od **r** članova, dopuštajući ponavljanje jednog te istog elementa A_k → svaki takav niz zovemo **varijacija n-tog reda i r-tog razreda s ponavljanjem**
- Koliko ima takvih varijacija?
  - 1. broj možemo odabrati na n načina
  - 2. broj možemo odabrati na n načina
  - r. broj možemo odabrati na n načina
- Slijedi da je broj varijacija n-tog reda i r-tog razreda s ponavljanjem:

$$\bar{V}_r^{(n)} = n^r$$

- Pri tome r može biti manji, jednak ili veći od n

Kombinacije bez ponavljanja

- Pretpostavimo da imamo **n različitih elemenata $A_1,\dots, A_n$** od kojih želimo **odabrati r** članova, ne dopuštajući ponavljanje jednog te istog elementa

  → Odabrana r-torka se naziva **kombinacija n-tog reda i r-tog razreda**
- Za razliku od varijacija ovdje ne pridajemo značaj poretku odabranih elemenata → dvije kombinacije razlikuju se samo ako sadrže različite elemente
- Koliko ima takvih kombinacija?
  - uzmimo da smo odabrali varijacije $V_r^{(n)}$
  - varijacije grupiramo u skupove koji sadrže iste elemente, ali s različitim poretkom → svaki takav skup će sadržavati r! varijacija
- Slijedi da je broj kombinacija:

$$K_r^{(n)} = \frac{V_r^{(n)}}{r!} = \frac{n!}{(n-r)!\,r!} \stackrel{\text{def}}{=} \binom{n}{r}$$

(čitamo: n povrh r)

Kombinacije bez ponavljanja

- Možemo postaviti dva ekvivalentna pitanja:
  - na koliko načina je moguće odabrati r elemenata od njih n? Odg. (n povrh r)
  - na koliko načina je moguće odabrati (n-r) elemenata od njih n? Odg. (n povrh n-r)
- Slijedi:

$$\binom{n}{n-r} = \binom{n}{r}$$

- Što se lako vidi i matematički:

$$\binom{n}{n-r} = \left(\frac{n!}{(n-r)![n-(n-r)]!}\right)\left(\frac{n!}{(n-r)!r!}\right) = \binom{n}{r}$$

Kombinacije bez ponavljanja

- Primjer loto 7/39:
  - Ukupno 39 brojeva
  - Treba odabrati 7 brojeva
    - ne ponavljaju se
    - poredak nije bitan

→ radi se o kombinacijama 39. reda i 7. razreda

→ K = 39!/(7! 32!) = 15 380 937

Statistika i osnovna mjerenja

Teorija vjerojatnosti

M. Makek

2016/2017

Uvod

- **Pokus** – bilo koji postupak ili proces koji rezultira opažanjem
- **Ishod** – moguć rezultat pokusa (različiti ishodi se međusobno isključuju)
- **Elementarni događaj** – pojedini ishod nekog pokusa
- **Prostor elementarnih događaja** – skup svih ishoda nekog pokusa
- Ako je Ω prostor elementarnih događaja onda je:
  - Događaj – svaki podskup od Ω
  - Elementarni događaj - jednočlani podskup od Ω
  - Složeni događaj – višečlani podskup od Ω
  - Sigurni događaj – cijeli Ω
  - Nemoguć događaj – prazan skup

Primjer: bacanje kocke

- 6 mogućih ishoda (elementarnih događaja)
- Složeni događaj – npr. parni ili neparni brojevi
- Siguran događaj – bilo koji broj od 1 - 6

Vjerojatnost a priori
(klasična definicija vjerojatnosti)

- Uzmimo pokus koji završava s konačno mnogo (*n*) ishoda tj. elementarnih događaja, koji su jednako mogući.
- Definicija: vjerojatnost proizvoljnog događaja *A* je dana omjerom broja *povoljnih* elementarnih događaja n_A i *ukupnog* broja elem. događaja *n*:

$$P(A) = \frac{n_A}{n}$$

- Iz ove definicije slijedi:
  - 0 <= P(A) <= 1 (vjerojatnost je između 0 i 1)
  - $n_A=0 \rightarrow P(A) = 0$ (ako niti jedan elementarni događaj ne realizira A, tj. A je nemoguć događaj, onda je njegova vjerojatnost jednaka 0)
  - $n_A=n \rightarrow P(A) = 1$ (ako svi elementarni događaji realiziraju A, tj. A je siguran događaj, njegova je vjerojatnost 1)
- Nedostaci ovakve definicije su:
  - pretpostavka o jednako mogućim događajima (tu već pretpostavljamo neku vjerojatnost pa je definicija kružna)
  - Definirana na konačnom skupu događaja

Vjerojatnost a priori

- Primjer 1: bacanje 2 kocke
  - Rezultat svakog bacanja je par brojeva (x_1,x_2)
  - Ukupno 36 elementarnih događaja čini prostor elementarnih događaja
  - Definirajmo ***složen događaj*** A kao onaj kod kojeg je suma brojeva na obje kocke jednaka 6
  - Elementarni događaji koji realiziraju A su:

    (1,5), (2,4),(3,3),(4,2),(5,1) → $n_A=6$

  → P(A) = 6/36 = 1/6

Vjerojatnost a priori

- Primjer 2: uzorak proizvoda
  - Pretpostavimo da imamo 500 proizvoda
  - Proizvođač garantira da je među njima maksimalno 5% neispravnih
  - Želimo ovo provjeriti uzimajući uzorak od 5 proizvoda

A. Koja je vjerojatnost da se među tih 5 proizvoda nađu 2 (tj. ~40%) loša?

1. Na koliko načina možemo odabrati uzorak od 5 proizvoda? → $K_5^{(500)} = \binom{500}{5}$
2. Na koliko načina se može odabrati uzorak u kojem je od 5 proizvoda 2 loša?

→ $K_2^{(25)} K_3^{(475)} = \binom{25}{2}\binom{475}{3}$

Vjerojatnost takvog događaja je: $P(A) = \frac{\binom{25}{2}\binom{475}{3}}{\binom{500}{5}}$ **= 0,021**

Koristimo tablicu logaritama faktorijela

Obzirom da je vjerojatnost takvog događaja vrlo mala (~2%) moglo bi se zaključiti da ako se među 5 nađu 2 loša proizvoda ukupni uzorak mora sadržavati više od 5% defektnih proizvoda i da garancija proizvođača nije ispunjena.

Za prethodni primjer:
Kako računati s velikim faktorijelama?

- Može se pojednostavniti:
  a) Korištenjem Stirlingove formule:

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

  b) Uporabom logaritama faktorijela
  (dano u tablicama)
  c) Logaritmiranjem Stirlingove formule a) + b)

Vjerojatnost a posteriori

- Primjer: bacanje novčića
  - Ako znamo daje novčić "pošten" onda možemo definirati ***a priori*** vjerojatnost:
    - Elementarni događaji P i G su jednako vjerojatni $p_P=p_G$
    - Slijedi: $p_P+p_G = 1 \rightarrow p_P=p_G=0{,}5$
  - Ako ne znamo da novčić daje jednako vjerojatne ishode onda možemo novčić bacati puno puta (n) i odrediti koliko puta se pojavljuje G (n_G), a koliko puta P (n_P)
  - Tada možemo definirati ***a posteriori*** vjerojatnost kao:

$$P(G) = \frac{n_G}{n} \qquad P(P) = \frac{n_P}{n}$$

Vjerojatnost a posteriori

- Definicija: vjerojatnost događaja **A** jednaka je relativnoj frekvenciji pojavljivanja tog događaja $f_r(A)$ u nizu od n pokusa:

$$f_r(A) = \frac{f_A}{n}$$

  - vrijedi pod pretpostavkom da su pokusi izvedeni pod jednakim uvjetima te da je izveden dovoljan broj pokusa da se $f_r(A)$ ne mijenja s n, tj. da su relativne frekvencije stabilne
- Nedostaci ovakve definicije:
  - Što znači dovoljan broj pokusa?
  - Koja je vjerojatnost jednog događaja (ako ne možemo ponavljati pokus puno puta)?

Suprotna vjerojatnost

- Vjerojatnost da se događaj *A* **ne** desi
  - Ne događanje ***A*** je događaj koji ćemo označiti sa $\bar{A}$
  - Ako je vjerojatnost za događaj *A, P(A),* onda je vjerojatnost da se *A* **ne desi**: $P(\bar{A}) = 1 - P(A)$
  - Suprotna vjerojatnost još se označava i sa Q(A) te vrijedi:

$$P(A) + Q(A) = 1$$

- Primjer: Imamo 10 proizvoda od kojih 5 dobrih i 5 neispravnih
  - Ako najednom uzmemo 3 proizvoda koja je vjerojatnost da barem jedan od njih bude neispravan – ekvivalentno: koja je vjerojatnost da se desi suprotan događaj onom kada su sva tri proizvoda ispravna?
  - Vjerojatnost da imamo 3 ispravna je: $P(A) = \binom{5}{3} / \binom{10}{3}$ =**1/12**
  - Vjerojatnost da nastupi suprotan događaj je:

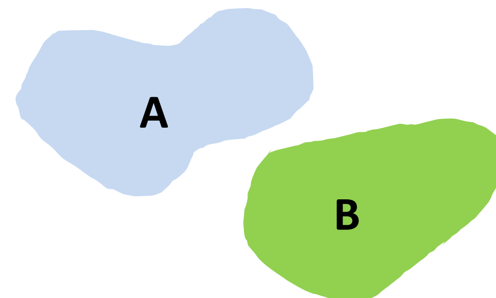
$$P(\bar{A}) = 1 - \frac{1}{12}$$ = **0,917**

Isključivost događaja

- Definicija: Događaji A i B se **međusobno isključuju** ako istovremeno ne mogu nastupiti oba.
  Analogno: događaji se međusobno isključuju ako se istovremeno može pojaviti samo jedan od njih

  → znači da ne postoji elementaran događaj koji bi istovremeno realizirao događaje A i B

  grafički se to može prikazati dis-junktnim skupovima točaka u ravnini



  → **skupovi elementarnih događaja koji realiziraju A odn. B su disjunktni**

Zbrajanje vjerojatnosti

- Pretpostavimo da se događaji $A_1, A_2, \ldots, A_k$ međusobno isključuju
- Zanima nas koja je vjerojatnost da se desi ili A_1 ili A_2, ..., ili A_k?
- Takav događaj označavamo s $A_1+A_2+\ldots+A_k$

  ("+" među događajima se čita "**ili**" i znači da se može desiti samo 1 događaj)

- Vrijedi: $$P\left(\sum_i A_i\right) = \sum_i P(A_i)$$

  vjerojatnost da se desi složeni događaj jednaka je sumi vjerojatnosti komponentnih događaja

- Dokaz slijedi iz definicije vjerojatnosti (npr. a priori):

$$P\left(\sum_i A_i\right) = \frac{\sum_i m_i}{n} = \sum_i \frac{m_i}{n} = \sum_i P(A_i)$$

Zbrajanje vjerojatnosti

- Primjer: bacanje kocke
  - Pretpostavimo da svakom broju na kocki pripada jednaka vjerojatnost P=1/6
  - Kolika je vjerojatnost da broj na kocki bude paran?

  → To je vjerojatnost da broj bude ili 2 ili 4 ili 6

  → Vjerojatnost za takav događaj je:

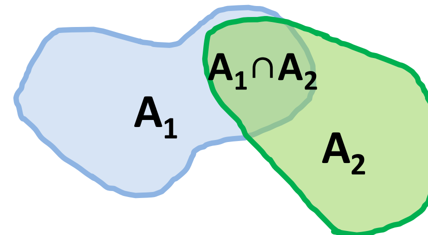
  P(Paran) = 1/6 + 1/6 + 1/6 = 0,5

Nezavisni događaji

- Događaji A_1 i A_2 su **nezavisni** ako pojavljivanje jednog ne mijenja vjerojatnost pojavljivanja drugog
- Kasnije ćemo pokazati da za nezavisne događaje vrijedi da je vjerojatnost istovremenog pojavljivanja događaja A_1 i A_2 jednaka umnošku vjerojatnosti pojedinih događaja, tj:

$$P(A_1 \cap A_2) = P(A_1)P(A_2)$$

- Grafički se može pokazati da je skup elementarnih događaja koji realiziraju i A_1 i A_2 jednak presjeku skupova elementarnih događaja za A_1 i A_2



- Implicitan uvjet je da se događaji A_1 i A_2 ne isključuju – drugim riječima događaji koji se isključuju ne mogu biti nezavisni

Množenje vjerojatnosti

- Pretpostavimo da se događaji $A_1, A_2, \ldots, A_k$ međusobno ne isključuju → mogu se dogoditi istovremeno
- Zanima nas koja je vjerojatnost da se istovremeno dese i A_1 i $A_2, \ldots,$ i A_k?
- Takav događaj označavamo s $A_1 \cap A_2 \cap \ldots \cap A_k$ ("∩" se čita "**i**")
- Može se pokazati da prethodni primjer vrijedi i općenito:

$$P\left(\bigcap_i A_i\right) = \prod_i P(A_i)$$

→ Vjerojatnost ovakvog složenog događaja je jednaka umnošku vjerojatnosti svih komponentnih događaja

Množenje vjerojatnosti

Primjer – koincidentno opažanje čestica

- Uzmimo da želimo opažati reakciju raspada piona:



$$\pi^0 \rightarrow e^+e^-\gamma$$

- U reakciji se kao produkti pojavljuju elektron, pozitron i foton
- Efikasnost detektora za opažanje elektrona ili pozitrona je ε_{e+e-}= 0,9
- Efikasnost detektora za opažanje fotona je ε_{γ}=0,2

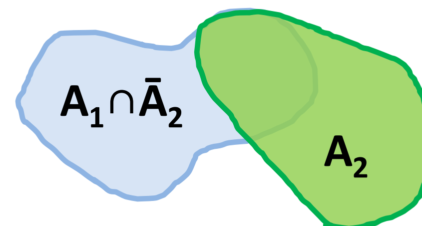
Kolika je vjerojatnost opažanja ovakvog događaja?

- Efikasnost detektora možemo poistovjetiti sa vjerojatnošću opažanja
- Ako pretpostavimo da je opažanje svake od čestica nezavisan događaj onda je vjerojatnost opažanja ovog događaja jednaka vjerojatnosti za istovremeno (koincidentno) opažanje ove tri čestice:

→ $P(\pi^0) = P(e^+)P(e^-)P(\gamma)$ = 0,162

Svojstva nezavisnih događaja

Tvrdnja: ako su A_1 i A_2 nezavisni onda su i A_1 i $\bar{A}_2$ nezavisni



- Iz grafičkog prikaza slijedi:

$$P(A_1 \cap \bar{A}_2) = P(A_1) - P(A_1 \cap A_2)$$

- Zbog nezavisnosti A_1 i A_2 možemo pisati:

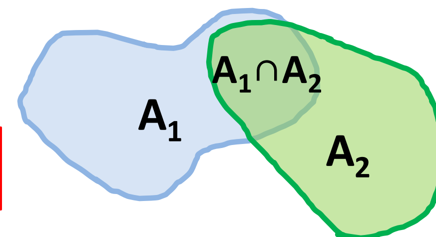
$$\begin{aligned} P(A_1 \cap \bar{A}_2) &= P(A_1) - P(A_1)P(A_2) \\ &= P(A_1)[1 - P(A_2)] \\ &= P(A_1)P(\bar{A}_2) \end{aligned}$$

- Čime je prema definicija nezavisnih događaja dokazana tvrdnja

Ostali složeni događaji

- Uzmimo događaje A_1 i A_2 koji su međusobno **ne** isključuju
- Kolika je vjerojatnost da nastupi barem jedan od tih događaja, tj. A_1 ili A_2 ili A_1 i A_2?
- Takav složeni događaj označavamo s $\boldsymbol{A_1} \cup \boldsymbol{A_2}$
- Iz grafičkog prikaza slijedi:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$



- U specijalnom slučaju kad su A_1 i A_2 nezavisni događaji, vrijedi:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1)P(A_2)$$

Ostali složeni događaji

- Prethodnu relaciju možemo napisati na drugačiji način:
  - po definiciji: $P(A_1 \cup A_2) = 1 - P(\overline{A_1 \cup A_2})$
  - $\overline{A_1 \cup A_2}$ znači da se **ne** dogodi A_1 ili A_2 ili oba
  - Ekvivalentno možemo reći da će se dogoditi događaj suprotan A_1 i suprotan A_2 : $\bar{A}_1 \cap \bar{A}_2$
- Slijedi: $P(A_1 \cup A_2) = 1 - P(\bar{A}_1 \cap \bar{A}_2)$
- Općenito vrijedi:

$$P(A_1 \cup A_2 \cup \ldots \cup A_k) = 1 - P(\bar{A}_1 \cap \bar{A}_2 \cap \ldots \cap \bar{A}_k)$$

- A posebno za nezavisne događaje:

$$P(A_1 \cup A_2 \cup \ldots \cup A_k) = 1 - P(\bar{A}_1)P(\bar{A}_2) \ldots P(\bar{A}_k)$$

Ostali složeni događaji

- <u>Primjer</u>:
  - A,B,C su nezavisni događaji koji se međusobno ne isključuju
  - U nekom pokusu oni imaju vjerojatnosti:
    P(A)=0.5, P(B)=0.3 i P(C)=0.1
  - Koja je vjerojatnost da nastupi bar jedan od događaja?

$$P(A \cup B \cup C) = 1 - P(\overline{A})P(\overline{B})P(\overline{C}) = 0{,}685$$

Konstantna vjerojatnost

- Događaji čija vjerojatnost se ne mijenja tijekom pokusa
  - Vjerojatnost da događaj nastupi u nekom pokusu:

$$P(A_1) = P(A_2) = \cdots = p$$

  - Vjerojatnost da događaj ne nastupi u nekom pokusu:

$$P(\bar{A}_1) = P(\bar{A}_2) = \cdots = 1 - p = q$$

  - Vjerojatnost da događaj nastupi barem jednom u seriji od k pokusa:

$$P(A_1 \cup A_2 \cup \ldots \cup A_k) = 1 - P(\bar{A}_1)P(\bar{A}_2)\ldots P(\bar{A}_k)$$

(događaji su po definiciji nezavisni jer se događaju u različitim pokusima)

  - Slijedi: $P(A_1 \cup A_2 \cup \ldots \cup A_k) = 1 - q^k$

Konstantna vjerojatnost

- <u>Primjer</u>:
  - U nekom pokusu događaj A nastupa uz vjerojatnost P(A)=p=0,4
  - Koliko pokusa treba učiniti da u njima sa vjerojatnošću od 95% nastupi barem jedan događaj?

    $P = 1-q^n = 0.95$

    →$0,6^n=0.05$

    → **n= 5,86 ~ 6**

Uvjetna vjerojatnost

- Kolika je vjerojatnost da se dogodi A ako se dogodio B?
  - Označava se sa **P(A|B)** (čitamo "A ako je B")
  - Skup elementarnih događaja koji realiziraju B onda postaje prostorom elementarnih događaja za A|B
  - U tom prostoru A∩B elementarnih događaja realizira dog. A
- Slijedi da je: $P(A|B) = \frac{P(A \cap B)}{P(B)}$ (uvjet P(B)>0)
- Jednako vrijedi: $P(B|A) = \frac{P(A \cap B)}{P(A)}$
- Iz čega slijedi: $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$

(vjerojatnost istovremenog zbivanja dva događaja jednaka je produktu apsolutne vjerojatnosti prvog i uvjetne vjerojatnosti drugog)

Nezavisnost događaja

- Definicija: događaji A i B su nezavisni onda i samo onda kad vrijedi:

  P (A|B) = P(A)

  (Vjerojatnost da se dogodi A ne ovisi o tome da li se dogodio B)

- Iz relacije za uvjetnu vjerojatnost vrijedi:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

- Iz toga slijedi relacija koju smo prije uveli:

$$\boxed{P(A \cap B) = P(A)P(B)}$$

Potpun sistem događaja

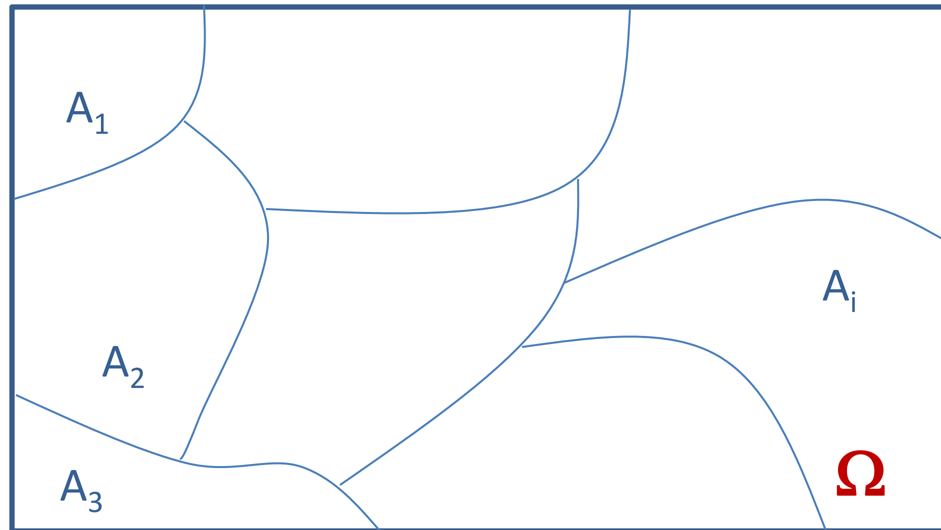
Neka za događaje A_i (i=1,2,...) vrijedi:

$$A_i \neq \emptyset, \forall i$$

$$A_i \cap A_j = \emptyset, \qquad \forall i, j, \qquad i \neq j$$

$$\bigcup_i A_i = \Omega$$

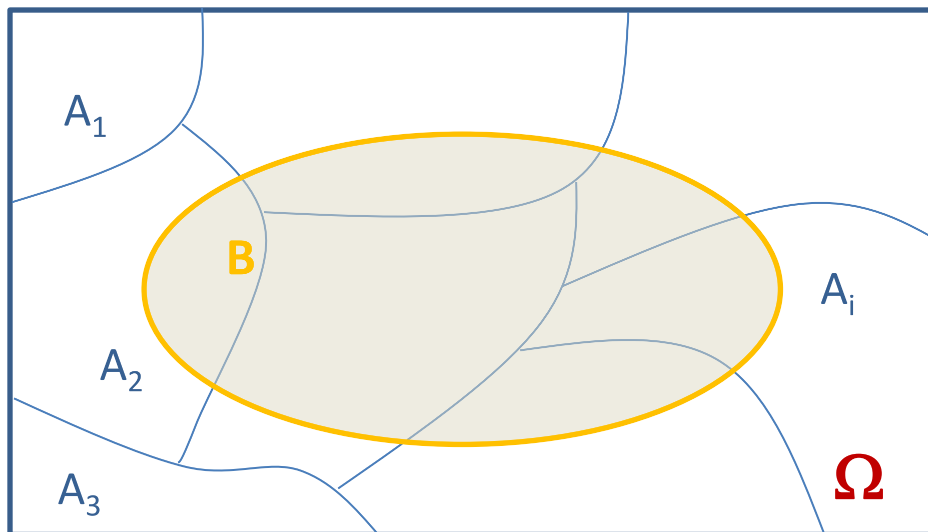
→ Događaji A_i čine potpun sistem događaja



Zakon totalne vjerojatnosti

- Neka A_i čine potpun sistem događaja
- Tada za bilo koji događaj B vrijedi:

$$P(B) = \sum_i P(B \cap A_i) \rightarrow P(B) = \sum_i P(B|A_i)P(A_i)$$



Bayesov teorem

- Neka A_i čine potpun sistem događaja i B neki događaj
- Za bilo koji događaj A_k vrijedi:

$$P(A_k|B) = \frac{P(A_k \cap B)}{P(B)}$$

- Slijedi ***Bayesov teorem***:

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_i P(B|A_i)P(A_i)}$$

Bayesov teorem

- Primjer: kontrola proizvoda
  - Serija proizvoda sadrži 95% ispravnih proizvoda
  - Kontrola proglašava ispravan proizvod dobrim s vjerojatnoću od 98%
  - Kontrola proglašava neispravan proizvod dobrim s vjerojatnošću od 5%
- Koja je vjerojatnost da je proizvod stvarno dobar ako ga je kontrola proglasila dobrim?
  - A_1 = proizvod stvarno dobar, $P(A_1) = 0.95$
  - A_2 = proizvod stvarno loš, $P(A_2) = 0.05$
  - B = kontrola proglašava proizvod dobrim
  - $P(A_1|B)$=?

Potpun sustav događaja (A_1, A_2)

→ $$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)} = 0{,}997$$

→ Nakon kontrole serija će sadržavati 99,7% dobrih proizvoda

Nedostaci klasičnih definicija vjerojatnosti

- A priori: vjerojatnost proizvoljnog događaja *A* je dana omjerom broja povoljnih elementarnih događaja n_A i ukupnog broja elem. događaja ***n***
- Nedostaci ovakve definicije su:
  - pretpostavka o jednako mogućim događajima (tu već pretpostavljamo neku vjerojatnost pa je definicija kružna)
  - Definirana na konačnom skupu događaja
- A posteriori: vjerojatnost događaja **A** jednaka je relativnoj frekvenciji pojavljivanja tog događaja $f_r(A)$ u nizu od ***n*** pokusa
- Nedostaci ovakve definicije:
  - Koliki je broj pokusa dovoljan?
  - Koja je vjerojatnost jednog događaja ?
(ako ne možemo ponavljati pokus puno puta)

Aksiomatska izgradnja teorije vjerojatnosti

- Ako poznajemo prostor elementarnih događaja Ω za neki pokus, svrha definicije vjerojatnosti je da svakom događaju $A \subseteq \Omega$ pridruži broj P(A), koji će biti precizna mjera šanse da se A ostvari
- Definicija vjerojatnosti treba biti općenita i obuhvaćati i klasične definicije vjerojatnosti
- Takvu definiciju temeljenu na aksiomima uveo je ***Kolmogorov*** 1933. godine
- Objekti u aksiomatskom pristupu su *slučajni događaji*

Kolmogorovi aksiomi

- Uzmimo da je Ω prostor elementarnih događaja
- Nekom događaju $A \subseteq \Omega$ želimo pridružiti vjerojatnost P(A)
- Pri tome tvrdimo da svako takvo pridruživanje mora zadovoljavati (aksiomi):

1. $P(A) \geq 0, \qquad \forall A$
2. $P(\Omega) = 1$
3. a) ako se konačan broj događaja međusobno isključuju (svaki par je disjunktan), vrijedi: $$P\left(\bigcup_{i}^{n} A_i\right) = \sum_{i}^{n} P(A_i)$$

   b) ako se prebrojivo beskonačan broj događaja međusobno isključuje, vrijedi: $$P\left(\bigcup_{i}^{\infty} A_i\right) = \sum_{i}^{\infty} P(A_i)$$

Svojstva vjerojatnosti

I.

a) Zbog zatvorenosti skupa elementarnih događaja: $A, \overline{A} \in \Omega$

b) Prema definiciji suprotnog događaja je: $A \cap \bar{A} = \emptyset$

c) Siguran događaj: $P(A \cup \bar{A}) = 1$

- Iz a), b), c) i trećeg Kolmogorovog aksioma slijedi

$$P(A) + P(\bar{A}) = 1$$

II.

Iz prvog aksioma i relacije (I.) slijedi: $0 \leq P(A) \leq 1$

III.

Ako u relaciju (I.) umjesto A uvrstimo siguran događaj Ω, dobivamo:

- Uz postavku da je $\overline{\Omega} = \emptyset$ slijedi: $P(\emptyset) = 0$

Svojstva vjerojatnosti

IV.

- Uzmimo da je $A_1 \subset A_2$
- Onda je $A_2 = A_1 \cup (A_2 - A_1)$
- Iz trećeg aksioma slijedi da je $P(A_2) = P(A_1) + P(A_2 - A_1)$
- Za bilo koja dva događaja vrijedi i:

$$A_1 \cup A_2 = A_1 \cup [A_2 - (A_1 \cap A_2)]$$

- Pa iz trećeg aksioma i gornjih relacija slijedi:

→ **$P(A \cup B) = P(A) + P(B) - P(A \cap B)$**

- Sva ova svojstva poznata su i u klasičnoj definiciji vjerojatnosti

Podudarnost sa klasičnom definicijom

- Neka pokus ima velik broj mogućih ishoda – elementarnih događaja ω_i
- Tada je vjerojatnost složenog događaja A: $P(A) = \sum_{\omega_i \subset A} P(\omega_i)$
- Ako imamo ***n*** jednako vjerojatnih ishoda (kao u definiciji a priori), onda je vjerojatnost svakog od njih: $P(\omega_i) = \frac{1}{n}, \forall i$
- Proizlazi da je vjerojatnost događaja A: $$P(A) = \frac{n_A}{n}$$

gdje je n_A broj elementarnih događaja koji su podskupovi A

→ to je klasična definicija vjerojatnosti!

Statistika i osnovna mjerenja

Slučajne varijable

M. Makek
2016/2017

Slučajna varijabla

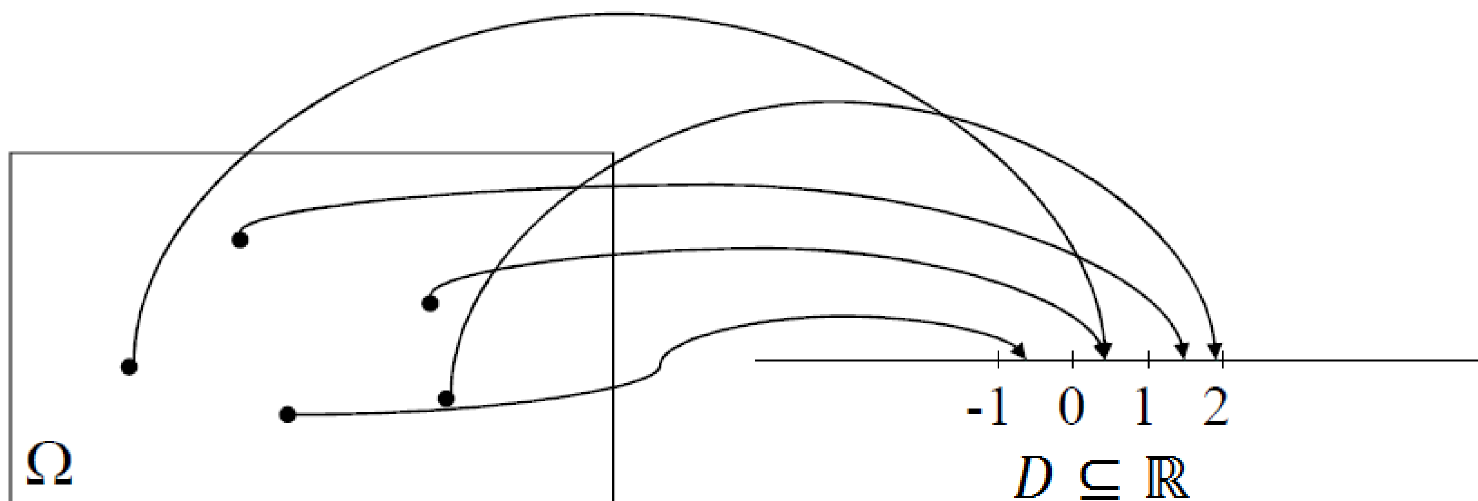
- Ω je prostor elementarnih događaja ω_i (definiranih u aksiomatskom, općenitom pristupu)
- ***slučajna varijabla X je funkcija,*** koja svakom ishodu pokusa ω_i iz Ω pridružuje neki broj ***x*** iz skupa realnih brojeva:

$$X : \Omega \to \mathbb{R}$$

- n.b. slučajna varijabla nije varijabla već funkcija i nije slučajna. Naziv potječe od činjenice da je domena od X skup slučajnih događaja, a obzirom da se uvriježio mi ćemo ga koristiti.

Slučajna varijabla

- Ako je Ω prostor elementarnih događaja slučajna varijabla pridružuje neki realni broj svakom elementarnom događaju ω.



D je skup svih vrijednosti slučajne varijable

Slučajna varijabla

<u>Primjer: bacanje 2 kocke</u>

- Prostor el. događaja Ω ima ukupno 36 elemenata $\omega=(i,j)$ gdje $i=1..6$ i $j=1..6$

a) Možemo definirati slučajnu varijablu X koja je jednaka sumi brojeva na kockama: $X(\omega)=i+j$

→ slučajna varijabla X događajima pridružuje vrijednosti 2…12 ili skraćeno kažemo slučajna varijabla poprima vrijednosti 2…12

b) Možemo definirati slučajnu varijablu Y koja je jednaka apsolutnoj razlici brojeva na kockama: $Y(\omega)=|i-j|$

→ slučajna varijabla Y događajima pridružuje vrijednosti 0…5 ili skraćeno kažemo slučajna varijabla poprima vrijednosti 0…5

Vrste slučajnih varijabli

- **Bernoullijeva slučajna varijabla** – s.v. koja opisuje događaj koji ima samo dva moguća ishoda (Bernoullijev događaj)
  - primjer: bacanje novčića
- **Diskretna slučajna varijabla** – s.v. za koju je skup mogućih vrijednosti konačan ili prebrojiv
  - primjer: bacanje novčića, bacanje kocke, rulet
- **Kontinuirana slučajna varijabla** – s.v. za koju je skup mogućih vrijednosti neki interval na brojevnom pravcu
  - primjer: vrijeme između 2 radioaktivna raspada iz nekog uzorka

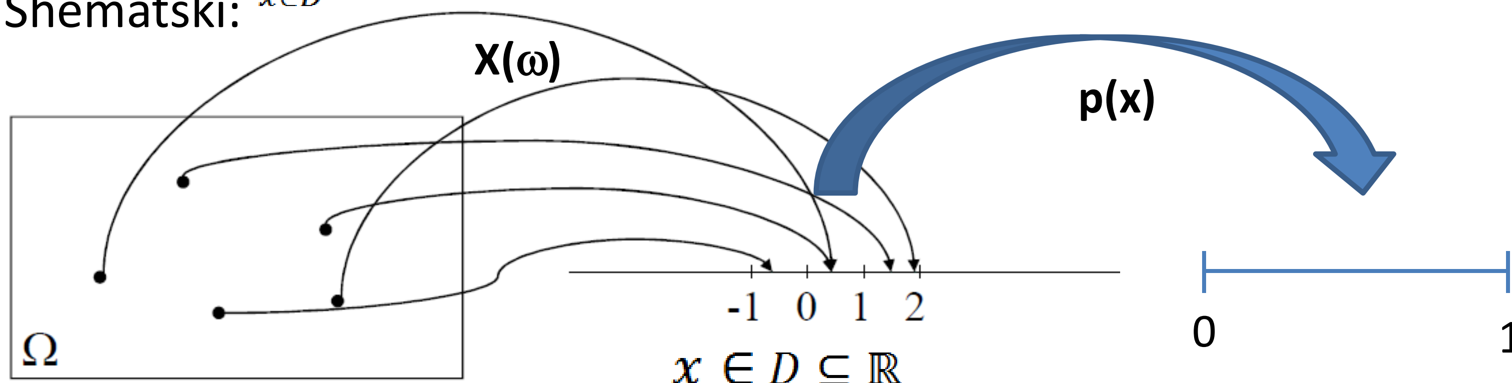
Diskretna slučajna varijabla

Def. *Raspodjela vjerojatnosti* diskretne slučajne varijable X definirana je za svaki $x \in D$:

$$p(x) = P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

(riječima: vjerojatnost da s.v. poprimi vrijednost x, koja odgovara vjerojatnosti da nastupi bilo koji elementarni događaj ω kojem je pridružena vrijednost x)

- Pri tome mora biti zadovoljen uvjet da suma svih vjerojatnosti mora biti jednaka 1: $\sum_{x \in D} p(x) = 1$
- Shematski:



Diskretna slučajna varijabla

Def. *kumulativna funkcija distribucije* (funkcija raspodjele) za diskretnu s.v. s raspodjelom vjerojatnosti *p(x)*:

$$F(x) = P(X \leq x) = \sum_{y \leq x} p(y)$$

(riječima: *F(x)* je vjerojatnost da s.v. poprimi vrijednost manju ili jednaku *x*)

- pri tome vrijedi: $F(-\infty) = 0 \; i \; F(+\infty) = 1$

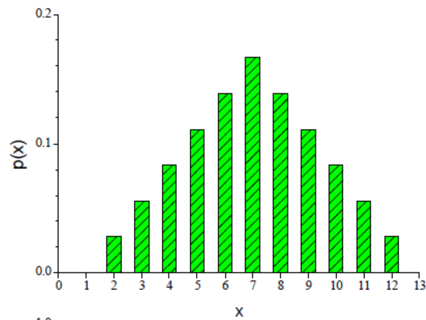
Za bilo koja dva broja a i b (a≤b) vrijedi: $P(a \leq X \leq b) = F(b) - F(a-)$

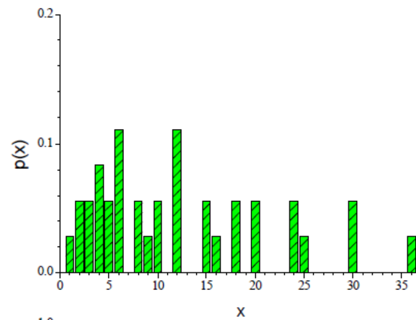
gdje je “a-” najveća vrijednost varijable X koja je < a.

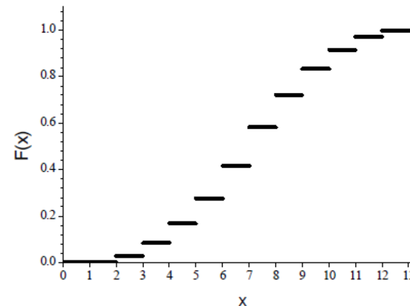
Diskretna slučajna varijabla
primjeri

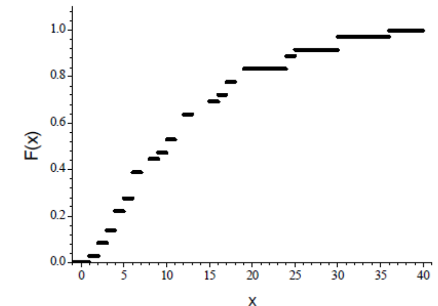
Raspodjele vjerojatnosti

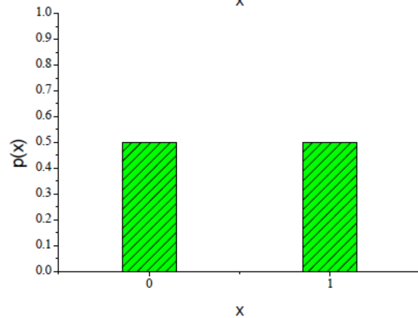
Kumulativne funkcije raspodjele

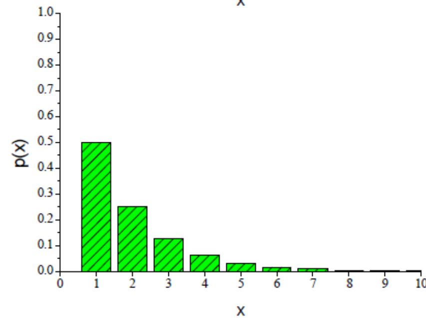


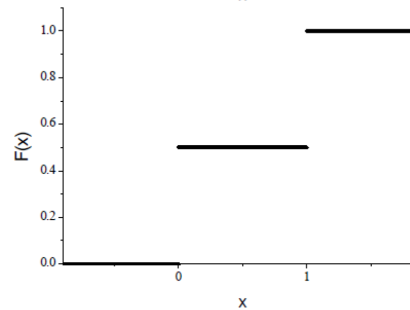


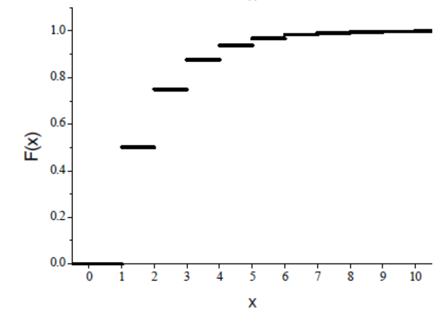












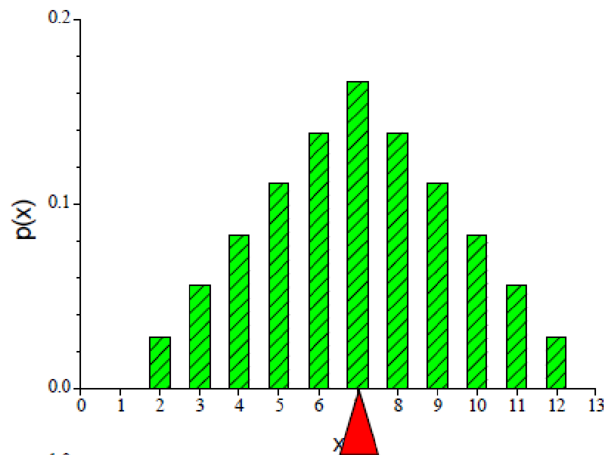
Diskretna slučajna varijabla

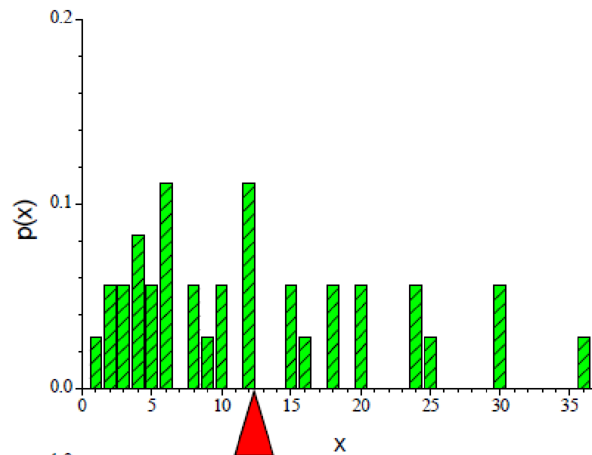
Def. *Očekivana vrijednost* diskretne slučajne varijable

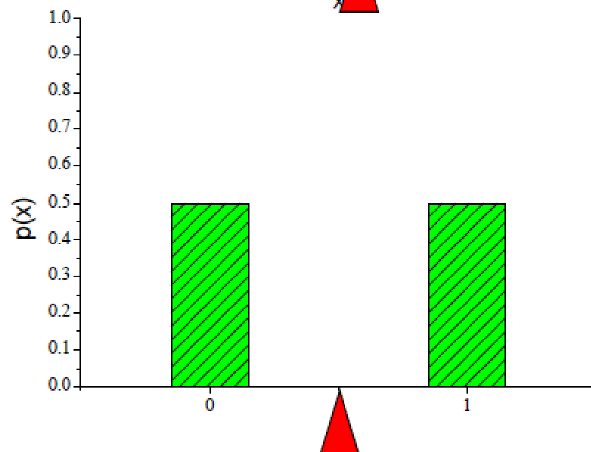
- Za diskretnu s.v. *X* sa skupom mogućih vrijednosti *D* i raspodjelom vjerojatnosti *p(x)* definirano očekivanje ili srednju vrijednost kao:

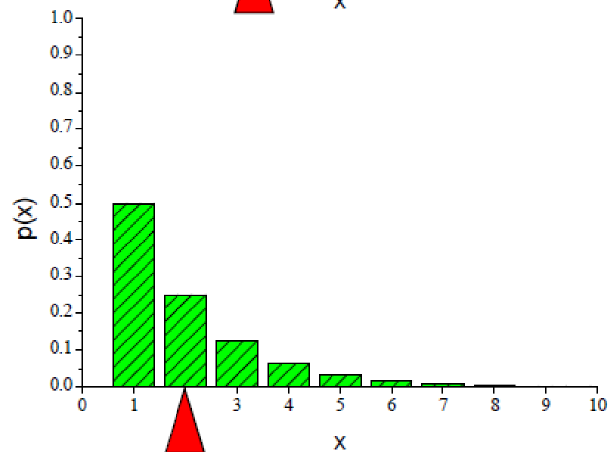
$$E(X) = \mu_X = \sum_{x \in D} x \cdot p(x)$$

Primjeri: očekivanje diskretne slučajne varijable









$$E(X) = \mu_X = \sum_{x \in D} x \cdot p(x)$$

Diskretna slučajna varijabla

Def. *Varijanca* diskretne slučajne varijable

- Za diskretnu s.v. *X* sa raspodjelom vjerojatnosti *p(x)* definira se varijanca:

$$V(X) = \sigma_X^2 = \sum_{x \in D} (x - \mu_X)^2 p(x) = E[(X - \mu_X)^2]$$

- Varijanca je jednaka sumi očekivanih kvadratnih odstupanja odn. jednaka je očekivanju kvadrata odstupanja od srednje vrijednosti
- $\sigma_X = \sqrt{V(X)}$ se naziva standardnom devijacijom ili standardnim odstupanjem s.v. *X*
- Također vrijedi: $V(X) = \sum_{x \in D} (x^2 - 2x\mu_X + \mu_x^2) p(x)$

$$V(X) = \sum_{x \in D} x^2 p(x) - 2\mu_X \sum_{x \in D} x p(x) + \mu_X^2 \sum_{x \in D} p(x)$$

$$V(X) = \sum_{x \in D} x^2 p(x) - 2\mu_X \mu_X + \mu_X^2 \quad \Rightarrow \quad V(X) = E(X^2) - [E(X)]^2$$

Primjer: bacanje 2 kocke

- Definiramo slučajnu varijablu *X* kao sumu ishoda bacanja 2 kocke
- X poprima vrijednosti: x=2…12
- Raspodjela vjerojatnosti:

x	2	3	4	5	6	7	8	9	10	11	12
p(x)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36
F(x)	1/36	3/36	6/36	10/36	15/36	21/36	26/36	30/36	33/36	35/36	36/36

- Očekivanje: *E(X)*= 7,0
- Standardna devijacija: σ_X = 2,4

Funkcija diskretne slučajne varijable

Očekivanje funkcije g(X) diskretne slučajne varijable *X je dano s*:

$$E\big(g(X)\big) = \sum_{x \in D} g(x)p(x)$$

- Specijalno za ***linearnu funkciju*** vrijedi:

$$E(aX + b) = \sum_{x \in D} (ax + b)p(x) = a \sum_{x \in D} xp(x) + b \sum_{x \in D} p(x)$$

  - Slijedi: $\boxed{E(aX + b) = aE(X) + b}$
  - U ovom specijalnom slučaju može se reći da je očekivanje funkcije jednako funkciji očekivanja

Funkcija diskretne slučajne varijable

- *Varijanca funkcije g(X)* diskretne slučajne varijable *X* je dana s:

$$V(g(x)) = \sum_{x \in D} \left(g(x) - E(g(X))\right)^2 p(x)$$

- može se pokazati da vrijedi:

$$V(g(X)) = E(g^2(X)) - [E(g(X))]^2$$

- Specijalno za ***linearnu funkciju*** vrijedi:

$$V(aX + b) = a^2 V(X)$$

  – Odnosno: $\sigma_{aX+b} = a\sigma_X$

Centralni moment r-tog reda

- Definira se kao:

$$M_r = E[(X-\mu)^r] = \sum_{x \in D} (x-\mu)^r p(x)$$

- Iz definicije slijedi:
  - ❑ $M_0 = 1$
  - ❑ $M_1 = 0$
  - ❑ $M_2 = \sum_{x \in D} (x-\mu)^2 p(x) = V(X)$

  → Centralni moment drugog reda jednak je varijanci

Pomoćni moment r-tog reda

- Definira se kao:

$$m_r = \sum_{x \in D} x^r p(x)$$

- Iz definicije slijedi:
  - ❑ $m_0 = 1$
  - ❑ $m_1 = E(X) = \mu_x$

→ pomoćni moment prvog reda jednak je očekivanju slučajne varijable

Veza centralnih i pomoćnih momenata

- Može se pokazati da su centralni i pomoćni momenti povezani relacijom (dokaz izostavljamo):

$$M_r = \sum_{k=1}^{r} (-1)^k \binom{r}{k} m_{r-k} m_1^k$$

- Npr.: $M_2 = m_2 - m_1^2$

  → Što je otprije poznata relacija $V(X) = E(X^2) - [E(X)]^2$

Funkcija izvodnica

- Za diskretnu slučajnu varijablu definiramo funkciju izvodnicu:

$$g(t) = \sum_{x \in D} e^{tx} p(x)$$

- Iz definicije slijedi:

$$g'(t = 0) = \sum_{x \in D} x \cdot p(x) = m_1$$

$$g''(t = 0) = \sum_{x \in D} x^2 \cdot p(x) = m_2$$

$$g^{(r)}(t = 0) = \sum_{x \in D} x^r \cdot p(x) = m_r$$

→ *pomoću funkcije izvodnice mogu se dobiti svi pomoćni, a posredno i centralni momenti slučajne varijable*

Momenti višeg reda

- Služe se detaljnu karakterizaciju raspodjela
- Moment trećeg reda:

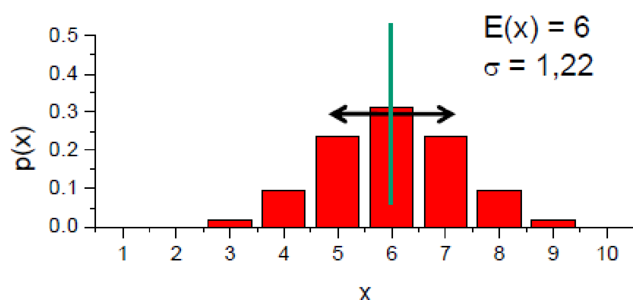
  → koeficijent asimetrije $\alpha_3 = \frac{M_3}{\sigma^3}$

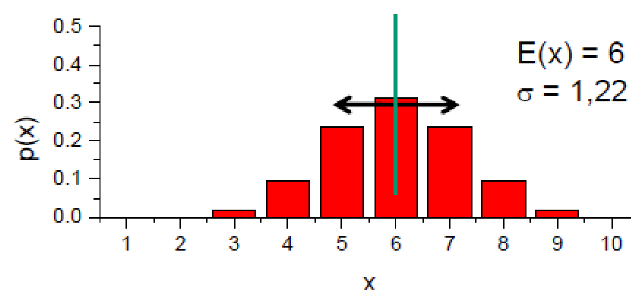
  - α_3 = simetrična raspodjela
  - $\alpha_3 > 0$ nagnuta udesno
  - $\alpha_3 < 0$ nagnuta ulijevo
- Moment četvrtog reda:

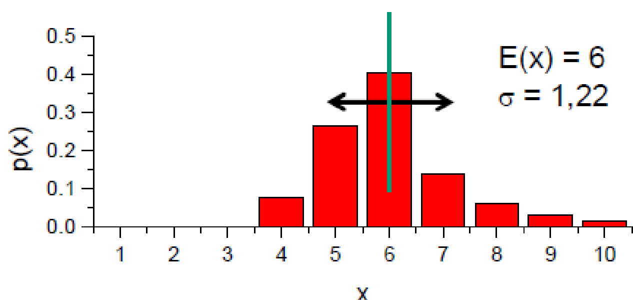
  → koeficijent spljoštenosti $\alpha_4 = \frac{M_4}{\sigma^4}$

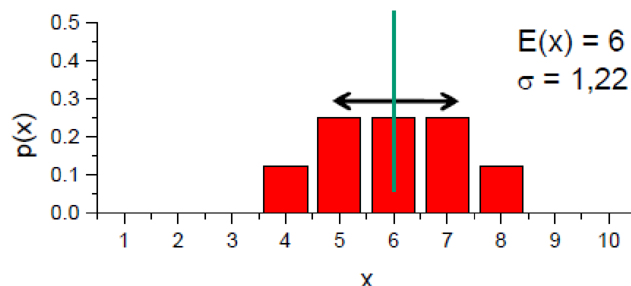
  - $\alpha_4 = 3$ normalno spoljoštena raspodjela
  - $\alpha_4 > 3$ šiljata raspodjela
  - $\alpha_4 < 3$ široka raspodjela

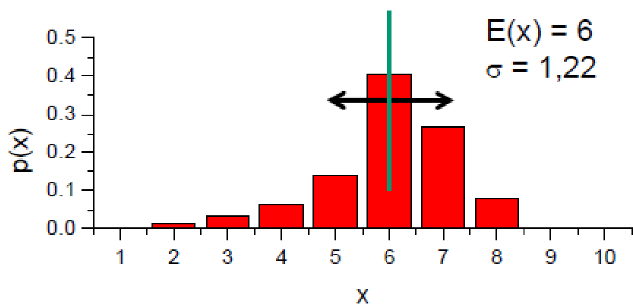
Momenti višeg reda

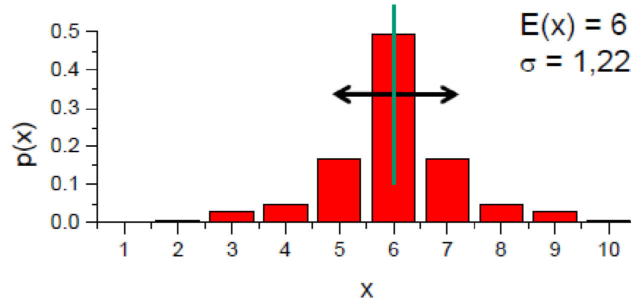






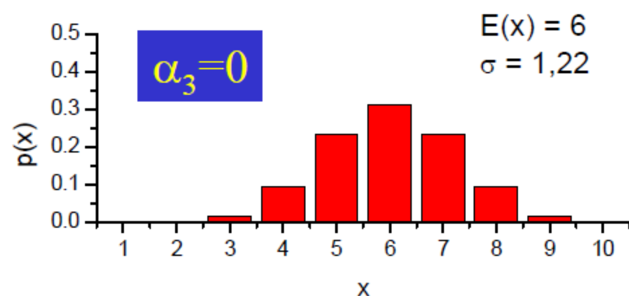


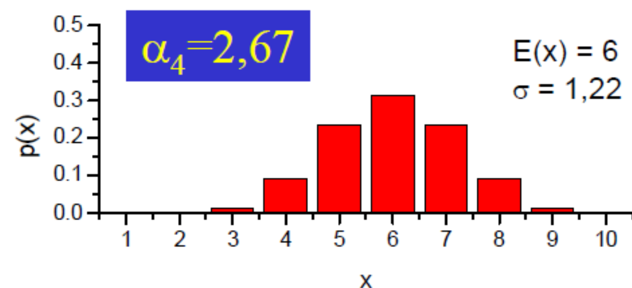


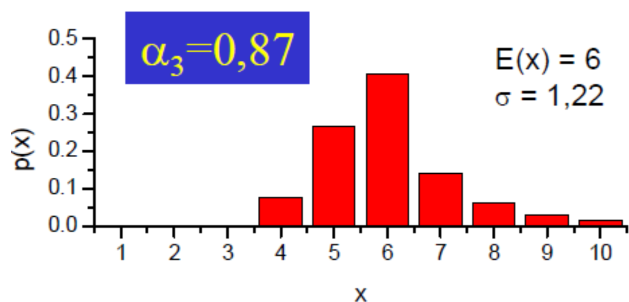


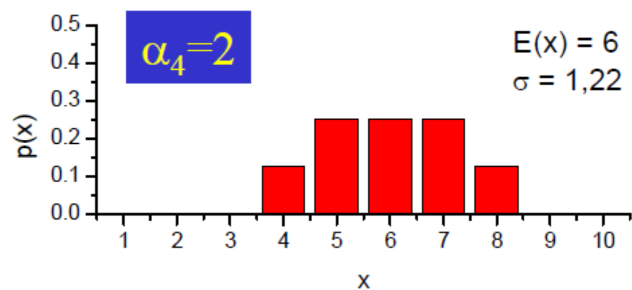
Sve raspodjele imaju jednako očekivanje i standardnu devijaciju
→ možemo ih bolje karakterizirati momoću momenata višeg reda

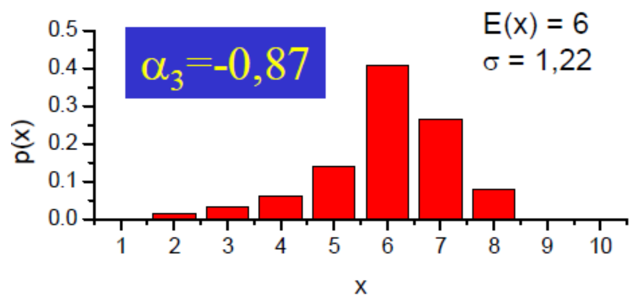
Momenti višeg reda

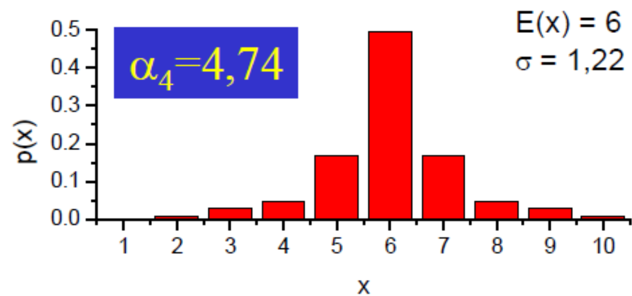












Statistika i osnovna mjerenja

Binomna i Poissonova raspodjela

M. Makek

2016/2017

Raspodjele vjerojatnosti

Želimo matematički i statistički opisati raspodjele varijabli koje se javljaju u prirodi u raznim područjima fizike i izvan nje

- Diskretne raspodjele: Bernoullijeva, Binomna, Poissonova, Hipergeometrijska, idr.
- Kontinuirane raspodjele: Gaussova, Gama, idr.

Binomna raspodjela

Primjeri:

- Bacanje novčića
  - Dva moguća ishoda P i G
  - Za pošten novčić p(P) = p(G) = 0.5
- Galtonova daska: http://www.geogebra.org/m/49479
  - Kuglica pada i na svakom koraku može skrenuti lijevo ili desno
  - p(L) = p(D) = 0.5

Što je zajedničko ovim pokusima?

→ Imaju dva moguća ishoda

- ➢ P ili G
- ➢ L ili D
- ➢ Uspjeh ili neuspjeh

Binomna raspodjela

Binomni pokus:

1. Sastoji se od ***n*** (Bernoullijevih) pokušaja, gdje je ***n*** unaprijed zadan
2. Pokušaji imaju dva moguća ishoda: uspjeh (***A***) ili neuspjeh ($\bar{A}$)
3. Vjerojatnost uspjeha (***p***) i neuspjeha (***q=1-p***) je jednaka za sve pokušaje
4. Pokušaji su nezavisni (ishod jednog ne utječe na drugi)

Binomna slučajna varijabla:

Za binomni pokus sa ***n*** pokušaja definiramo slučajnu varijablu:

X = broj uspjeha u ***n*** pokušaja, **X~Bin(n,p)**

X = 0…n

Binomna raspodjela

- Def. Binomna raspodjela
  - Za binomni eksperiment i binomnu slučajnu varijablu ***X~Bin(n,p)***, njenu raspodjelu vjerojatnosti nazivamo binomnom raspodjelom i označavamo: ***b(x; n,p)***
- Uzmimo da niz od n Bernoullijevih eksperimenata završi:

  $AA\bar{A}A\bar{A}A\ldots\ldots\ \bar{A}A$
  - A se pojavljuje x puta, $\bar{A}$ n-x puta
  - Koja je vjerojatnost jednog takvog ishoda? → $p^x q^{n-x}$
  - Koliko ima takvih istih ishoda (koji su svi jednako vjerojatni)? → $\binom{n}{x}$
  - Vjerojatnost da se pojavi bilo koji ishod koji ima **x** uspjeha je:

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x}$$

za x=0,1,2,...n

b(x; n, p) = 0 za sve ostale x

→ vjerojatnost da u seriji od ***n*** pokusa događaj ***A*** nastupi ***x*** puta

Binomna raspodjela

- <u>Primjer</u>: Galtonova daska

$$b(x;4{,}0.5)=\binom{4}{x}0.5^{x}0.5^{4-x}$$

$$b(x;4{,}0.5)=\binom{4}{x}\frac{1}{16}$$

slučajna varijabla x	putevi	vjerojatnost
0	LLLL	q^4
1	LLLD LLDL LDLL DLLL	$p \cdot q^3$
2	LLDD LDLD LDDL DLLD DLDL DDLL	$p^2 \cdot q^2$
3	LDDD DLDD DDLD DDDL	$p^3 \cdot q$
4	DDDD	p^4

Binomna raspodjela

Rekurzivna formula

- Pomaže u računu, posebice kod velikih faktorijela
- Koja je vjerojatnost da varijabla poprimi vrijednost x-1?

$$b(x-1;n,p) = \binom{n}{x-1} p^{x-1} q^{n-x+1}$$

- Iz omjera **b(x)/b(x-1)** dobivamo rekurzivnu formulu:

$$b(x;n,p) = \frac{n-x+1}{x}\frac{p}{q}b(x-1;n,p)$$

→ Vjerojatnost za bilo koji x možemo dobiti rekurzijom iz:

$$b(0;n,p) = q^n$$

Binomna raspodjela

Primjer korištenja rekurzije: za koju vrijednost x je b(x) maksimalan?

- Koristimo rekurzivnu formulu

Mora biti zadovoljeno:

$$b(x-1;n,p) \leq b(x;n,p) \geq b(x+1;n,p)$$

Nakon raspisivanja (domaća zadaća) po rekurzivnoj formuli slijedi:

$$np - q \leq x_{max} \leq np + p$$

Binomna raspodjela

- Vrijedi binomni poučak:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Funkcija raspodjele diskretne slučajne varijable sa binomnom raspodjelom:

$$F(x) = \begin{cases} \sum_{k=0}^{x} \binom{n}{k} p^k q^{n-k}, & 0 \leq x < n \\ 1, & x = n \end{cases}$$

Binomna raspodjela

Očekivanje diskretne slučajne varijable sa binomnom raspodjelom je: $E(X) = np$

- To možemo pokazati na sljedeći način:
  - Definiramo funkciju $g(t) = (q + pt)^n$
  - Razvoj po binomnom poučku:

$$(q + pt)^n = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} t^x = \sum_{x=0}^{n} b(x) t^x$$

  - Deriviranjem po t dobivamo:

$$np(q + pt)^{n-1} = \sum_{x=0}^{n} x\, b(x) t^{x-1}$$

  - Za t=1:

$$np(q + p)^{n-1} = \sum_{x=0}^{n} x\, b(x)$$

  - Što po definiciji daje:

$$np = E(X)$$

Funkcija izvodnica

- Općenito za diskretnu slučajnu varijablu definiramo funkciju izvodnicu:

$$g(t) = \sum_{x \in D} e^{tx} p(x)$$

- Iz definicije slijedi:

$$g'(t=0) = \sum_{x \in D} x \cdot p(x) = m_1$$

$$g''(t=0) = \sum_{x \in D} x^2 \cdot p(x) = m_2$$

$$g^{(r)}(t=0) = \sum_{x \in D} x^r \cdot p(x) = m_r$$

→ *pomoću funkcije izvodnice mogu se dobiti svi pomoćni, a posredno i centralni momenti slučajne varijable*

Binomna raspodjela

- Drugi način da odredimo očekivanje (a i varijancu) je pomoću funkcije izvodnice
- Funkcija izvodnica za binomnu raspodjelu:

$$g_X(t) = E(e^{tX}) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x q^{n-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x}$$

- Prema binomnom poučku slijedi:

$$g_X(t) = (pe^t + q)^n$$

Binomna raspodjela

- Prva derivacija funkcije izvodnice:

$$g'_X(t) = npe^t(pe^t + q)^{n-1}$$

$$g'_X(0) = npe^0(pe^0 + q)^{n-1} = np \qquad \textbf{(=} \; \boldsymbol{m_1}\textbf{)}$$

$$\Rightarrow \boxed{E(X) = np}$$

- Druga derivacija funkcije izvodnice:

$$g''_X(t) = npe^t(pe^t + q)^{n-1} + n(n-1)(pe^t)^2(pe^t + q)^{n-2}$$

$$g''_X(0) = np + n(n-1)p^2 \qquad \textbf{(=} \; \boldsymbol{m_2}\textbf{)}$$

Binomna raspodjela

- Varijanca je jednaka centralnom momentu drugog reda
- Otprije imamo: $M_2 = m_2 - m_1^2$
- Slijedi Varijanca diskretne slučajne varijable sa binomnom raspodjelom:

$$V(X) = np + n(n-1)p^2 - (np)^2$$
$$= np(1 + (n-1)p - np)$$
$$= np(1-p)$$

$$\Rightarrow \boxed{V(X) = npq}$$

Treba još primijetiti da se za n=1 binomna raspodjela svodi na Bernoullijevu, što je i za očekivati budući da ona opisuje Bernoullijev pokus koji se ponavlja n puta!

Binomna raspodjela

Primjer: kontrola uzoraka

- Neki uređaj izrađuje 8% neispravnih proizvoda. Proizvodi se pakiraju u kutije od 50 komada.

  Koja je vjerojatnost da se u kutiji nađe 2 defektna komada?

  - n=50, p=0.08
  - $b(2) = \binom{50}{2} p^2 q^{50-2}$

    → **b(2) = 0.14**

  Koji je najvjerojatniji broj defektnih komada u kutiji?

  - n=50, p=0.08
  - np-q <= x_M <= np+p → 3.08 < = x_M <= 4.08

    → **x_M = 4**

Poissonova raspodjela

- Uzmimo da je vjerojatnost uspjeha pojedinog Bernoullijevog pokušaja u binomnom pokusu jako mala, te taj pokus ponavljamo jako puno puta (**n >>, p <<**)

  pri tome očekivana vrijednost konstantna **E(x)=np= const.**

- Primjer: raspad radioaktivnog uzorka:
  - Vjerojatnost pojedinog raspada je iznimno mala
  - Uzorak sadrži jako puno izotopa koji se nezavisno mogu raspasti

  → Srednji broj raspada u nekom vremenskom intervalu je konstantan.

Poissonova raspodjela

- Kao granični slučaj binomne:

$$b(x;n,p) = \binom{n}{x} p^x q^{n-x}$$

$$= \frac{n(n-1)\ldots(n-x+1)}{x!} p^x q^{n-x}$$

$$= \frac{n(n-1)(n-x+1)}{x!} \frac{\lambda^x}{n^x} \left(1-\frac{\lambda}{n}\right)^{n-x}$$

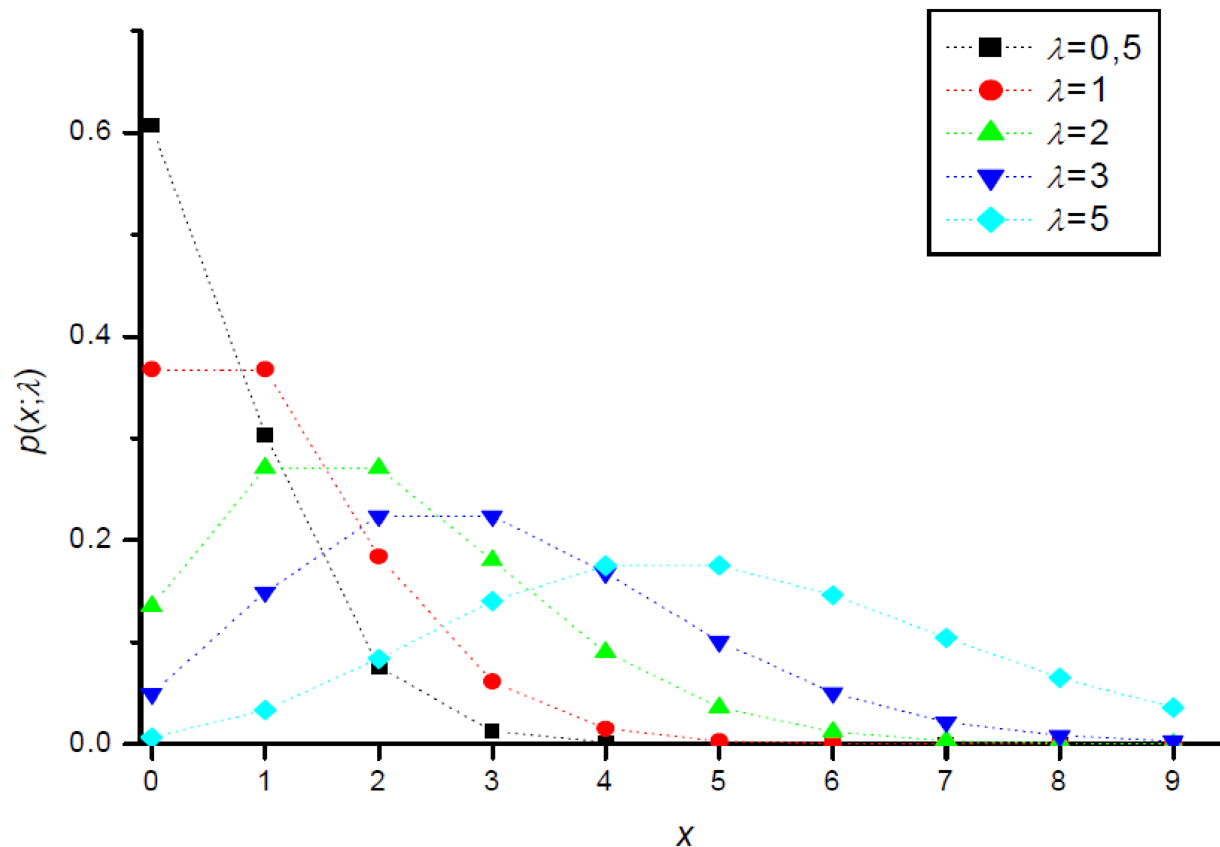
$$= \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{x-1}{n}\right)}{x!} \lambda^x \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x}$$

$$p(x) = \lim_{n\to\infty} b(x) = \frac{\lambda^x}{x!} \lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^n$$

$$\Rightarrow \boxed{p(x) = \frac{\lambda^x}{x!} e^{-\lambda}}$$

Poissonova raspodjela

- P.r. je potpuno određena parametrom λ



Poissonova raspodjela

- Očekivanje je prema definiciji:

$$E(X) = \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda}$$

$$= \lambda e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} e^{\lambda}$$

$$\Rightarrow \boxed{E(X) = \lambda}$$

- Prema definiciji varijance također izlazi da je:

$$\boxed{V(X) = \lambda} \text{ odn. } \boxed{\sigma_X = \sqrt{\lambda}}$$

- Normiranost raspodjele slijedi iz definicije:

$$\sum_{x=0}^{\infty} p(x;\lambda) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$$

Poissonova raspodjela

- Funkcija izvodnica:

$$g(t) = \sum_{x=0}^{\infty} e^{tx} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

- Pomoćni momenti:

$$g'(t) = \lambda e^t e^{\lambda(e^t-1)} \rightarrow \; m_1 = \lambda$$

$$g''(t) = \lambda e^t (1 + \lambda e^t) e^{\lambda(e^t-1)} \rightarrow m_2 = \lambda^2 + \lambda$$

→ ***E(X)*** i ***V(X)*** možemo elegantno dobiti preko funkcije izvodnice

Poissonova raspodjela

- Rekurzija
  - Iz omjera p(x)/p(x-1) slijedi:

$$p(x) = \frac{\lambda}{x} p(x-1)$$

- Najvjerojatnija vrijednost
  - Iz uvjeta: $p(x_M - 1) \leq p(x_M) \geq p(x_M + 1)$
  - Slijedi: $\lambda - 1 \leq x_M \leq \lambda$

Raspodjela dvije nezavisne Poissonove varijable

- Za dvije slučajne varijable ***X*** i ***Y*** s Poissonovim raspodjelama ***p(x,m)*** i ***p(y,n)*** vrijedi da je raspodjela slučajne varijable ***Z=X+Y*** jednaka ***p(z, m+n)***
- Dokaz:

$$P(Z = z) = \sum_{k=0}^{z} P(X = k)P(Y = z - k)$$

$$= \sum_{k=0}^{z} e^{-m} \frac{m^k}{k!} \, e^{-n} \frac{n^{z-k}}{(z-k)!}$$

$$= \frac{e^{-(m+n)}}{z!} \sum_{k=0}^{z} \binom{z}{k} m^k n^{z-k}$$

$$= \frac{e^{-(m+n)}}{z!} (m+n)^z = p(z, m+n)$$

Poissonova raspodjela
(primjene)

Primjer: **Poissonova raspodjela kao granični slučaj binomne**

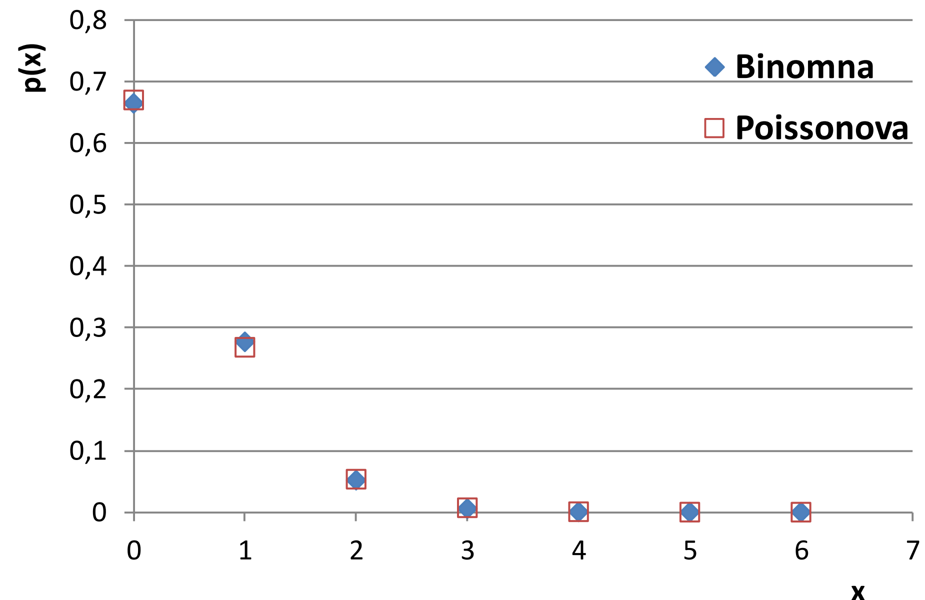
- Serija proizvoda sadrži 4% loših
- Kolika je vjerojatnost da u seriji od 10 proizvoda nadjemo najviše jedan loš proizvod?
  - N=10
  - P=0.04

  b(0;10,0.04) + b(1;10,0.04) = 0.9418

  - Np = 0.4

  p(0;0.4) + p(1;0.4) = 0.9384

→ Poissonova r. je dobra aproksimacija binomne!



Poissonova raspodjela
(primjene)

Primjer: **Poissonova raspodjela za proučavanje rijetkih događaja**

- Konstanta raspada nekog izotopa je λ=7,3 10^{-10} s^{-1}
- Uzorak sadrži 4,8 10^{22} radioaktivnih atoma
- Treba odrediti raspodjelu vjerojatnosti ***k*** raspada u 1 ms.
- Zakon radioaktivnog raspada daje broj atoma u vremenu:

$$N(t) = N_0 e^{-t/\tau}$$

- I aktivnost uzorka: $a = \left|\frac{dN}{dt}\right| = \frac{N}{\tau}$
- Slijedi da je vjerojatnost k raspada u intevalu Δt:

$$P_k(\Delta t) = b(k; N, \frac{N\Delta t}{\tau}\frac{1}{N}) = p(k; a\Delta t)$$

- Vjerojatnost događaja proporcionalna je ***veličini intervala***

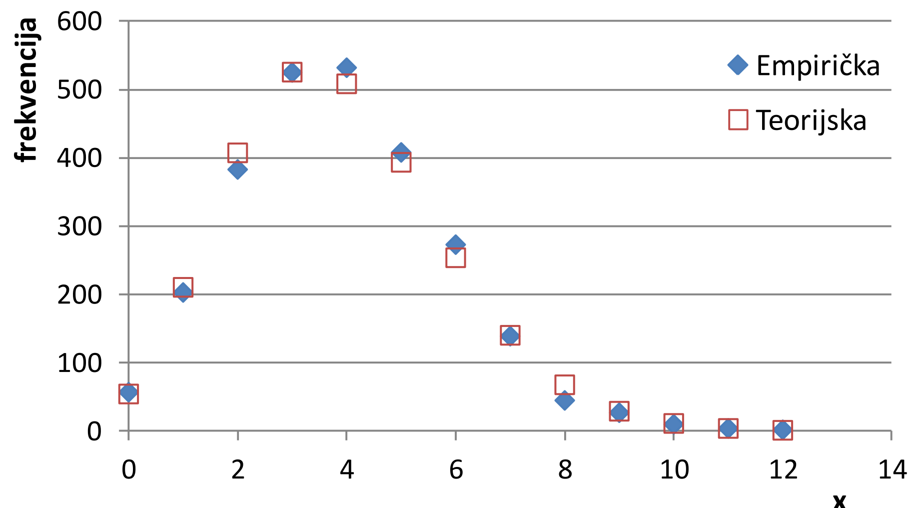
→ Takve procese nazivamo Poissonovim procesima!

Prilagodba empiričkim podacima

- Primjer Poissonova raspodjela
- Mjeren je broj beta raspada ***x*** iz nekog uzorka u intervalu od 10 sekundi. Broj slučajeva u kojima je emitirano točno ***x*** čestica je označen kao $\boldsymbol{f_x}$ → smatramo da bi broj opaženih čestica trebao slijediti Poiss. raspodjelu.

x	f_x	p(x)	f_{tx}
0	57	0,0209	54,4
1	203	0,0807	210,5
2	383	0,1562	407,4
3	525	0,2015	525,5
4	532	0,1949	508,4
5	408	0,1509	393,5
6	273	0,0973	253,8
7	139	0,0538	140,3
8	45	0,0260	67,9
9	27	0,0112	29,2
10	10	0,0043	11,3
11	4	0,0015	4,0
12	2	0,0005	1,3

- Očekivanje raspodjele možemo aproksimirati aritmetičkom sredinom uzorka: $\lambda = \bar{x}$ =3,87
- Teorijske frekvencije računamo prema: $f_{tx} = Np(x;\lambda)$



Druge raspodjele diskretnih varijabli

- **Hipergeometrijska raspodjela**
  - Populacija od N elemenata (konačna populacija)
  - Svaki element može biti uspješan (A) ili neuspješan (Ā)
  - Postoji M uspješnih elemenata
  - Izvlačenje bilo kojeg uzorka od ***n*** elemenata je jednako vjerojatno
  - Slučajna varijabla ***X=broj uspješnih elemenata u uzorku***
  - **h(x)**=broj mogućih uzoraka s ***x*** uspjeha/ukupni broj mogućih uzoraka

$$h(x;n,M,N)=\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \qquad x=0,1,2,\ldots,\min(n,M)$$

Druge raspodjele diskretnih varijabli

- **Hipergeometrijska raspodjela**
  - Očekivanje $E(X) = n\frac{M}{N}$
  - Varijanca $V(X) = n\frac{M}{N}\frac{N-M}{N}\frac{N-n}{N-1}$
  - Uz definicije: p=M/N i q=(N-M)/N slijedi:

$$E(X) = np \qquad V(X) = npq\frac{N-n}{N-1}$$

- Primjer: skup ima 50 proizvoda od kojih 40 dobrih i 10 loših. Slučajno odaberemo 5 proizvoda. Koja je vjerojatnost da se u uzorku nađe x dobrih proizvoda? -> h(x; 5,40,50)

Druge raspodjele diskretnih varijabli

- **Pascalova raspodjela**
  - Niz nezavnisih pokušaja
  - Svaki pokušaj je Bernoullijev pokus koji može biti uspješan (A) ili neuspješan (Ā)
  - Vjerojatnost uspjeha ***p*** jednaka je za sve pokušaje
  - Pokušaji se izvode sve dok se ne opazi ***r*** uspjeha (r unaprijed određen)
  - Slučajna varijabla ***X = broj neuspjeha koji prethode r-tom uspjehu***
  - Za razliku od binomne broj pokušaja je varijabilan, a broj uspjeha je zadan

$$nb(x;r,p) = P(r-1\,A\,u\,x+r-1\,pokušaja)P(uspjeh)$$

$$nb(x;r,p) = \binom{x+r-1}{r-1} p^{r-1} q^x p$$

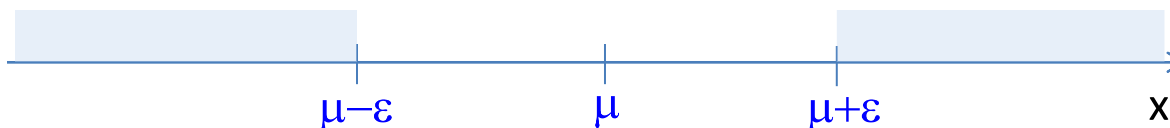
$$nb(x;r,p) = \binom{x+r-1}{r-1} p^r q^x \qquad x = 0,1,2 \ldots$$

Čebiševljev teorem

- Za slučajnu varijablu ***X*** sa raspodjelom vjerojatnosti ***P(x)*** i očekivanjem ***E(X)=μ*** i varijancom ***V(X)=σ^2*** vrijedi teorem:

  Vjerojatnost da ***X*** poprimi vrijednost izvan intervala **(μ–ε, μ+ε)** je manja ili jednaka **σ^2/ε^2**, za bilo koji **ε>0**.

$$P\{|x-\mu| \geq \epsilon\} \leq \frac{\sigma^2}{\epsilon^2}$$



Dokaz: po definiciji je varijanca jednaka: $\sigma^2 = \sum_i (x_i - \mu)^2 P(x_i)$

Posebno je za sve x iz osjenčanog intervala:

$$\sigma^2 \geq \sum_{|x_i-\mu|\geq\epsilon} (x_i - \mu)^2 P(x_i)$$

Obzirom da je: $|x_i - \mu| \geq \epsilon$ onda vrijedi:

$$\sigma^2 \geq \epsilon^2 \sum_{|x_i-\mu|\geq\epsilon} P(x_i)$$

Budući da se x_i međusobno isključuju vrijedi: $\sum_{|x_i-\mu|\geq\epsilon} P(x_i) = P\{|x-\mu| \geq \epsilon\}$

Odavde slijedi tvrdnja

Zakon velikih brojeva

- Za Bernoullijev događaj s vjerojatnošću ***P(A)=p***, koji se ponavlja u ***n*** pokusa, vjerojatnost slijedi binomnu raspodjelu sa očekivanjem $\mu = np$ i varijancom $\sigma^2 = npq$. Empirijski je relativna frekvencija tog događaja: $f_{rx} = \frac{x}{n}$ (gdje je x broj nastupa u n pokušaja)

Bernoullijev teorem: ako broj pokušaja teži u beskonačno onda relativna frekvencija teži *po vjerojatnosti* stvarnoj vjerojatnosti događaja:

$$\lim_{n \to \infty} P\{|f_{rx} - p| \geq \epsilon\} = 0$$ (za bilo koji ε)

<u>Dokaz</u>

$$P\{|f_{rx} - p| \geq \epsilon\} = P\left\{\left|\frac{x}{n} - p\right| \geq \epsilon\right\} = P\{|x - np| \geq n\epsilon\}$$

Prema Čeb. Tm (uz np=μ): $P\{|x - np| \geq n\epsilon\} \leq \frac{\sigma^2}{n^2\epsilon^2}$

$$P\{|f_{rx} - p| \geq \epsilon\} \leq \frac{\sigma^2}{n^2\epsilon^2} = \frac{npq}{n^2\epsilon^2} \Rightarrow P\{|f_{rx} - p| \geq \epsilon\} \leq \frac{pq}{n\epsilon^2}$$

Za lim(n-> beskonačno:) $P\{|f_{rx} - p| \geq \epsilon\} = 0$ (Q.E.D.)

Statistika i osnovna mjerenja

Kontinuirane raspodjele

M. Makek
2016/2017

Kontinuirana slučajna varijabla

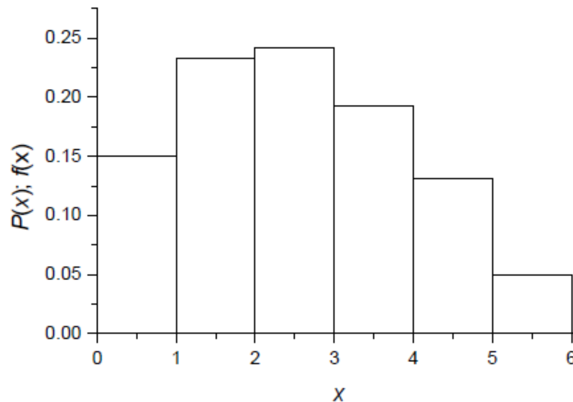
Definicija: Slučajna varijabla ***X*** je **kontinuirana** ako skup njenih vrijednosti čini interval brojeva, tj. ako je za ***A*** i ***B*** **(*A*<*B*)** svaki ***x* Є [*A*,*B*]** moguć.

- Kontinuirana slučajna varijabla će biti reprezentacija opservabli poput mase, duljine, vremena, itd.
- Većina karakterističnih veličina (očekivanje, varijanca, idr.) kontinuirane slučajne varijable može se dobiti iz veličina diskretne slučajne varijable uz prelazak sa ***sume*** na ***integral***

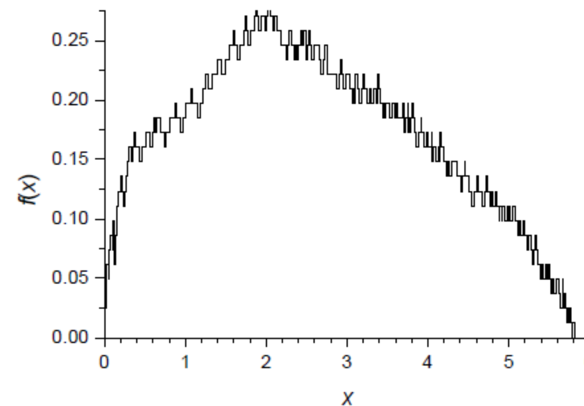
Kontinuirana slučajna varijabla

- Primjer: mjerimo dubinu jezera na ~10000 mjesta

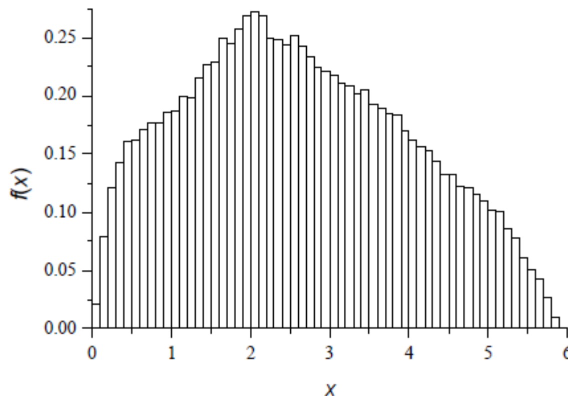
zaokruženo na 1 m



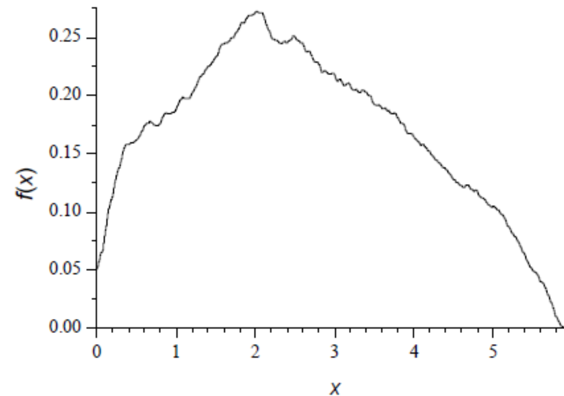
zaokruženo na 1 cm



zaokruženo na 1 dm



kontinuirana raspodjela



Prijelaz s diskretne na kontinuiranu: $p(x) \to f(x) \cdot dx$ $\sum_{x \in D} \to \int_{-\infty}^{\infty}$

Funkcija gustoće vjerojatnosti

Definicija: ***funkcija gustoće vjerojatnosti*** kontinuirane slučajne varijable ***X*** je funkcija ***f(x)*** takva da za bilo koja dva broja ***a*** ≤ ***b*** vrijedi:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Iz definicije slijedi da je:
$P(X=x_0) = 0$, za svaki x_0!

→ *Vjerojatnost da X poprimi vrijednost u intervalu [a,b] dana je površinom ispod funkcije gustoće vjerojatnosti u tom intervalu*

- Pri tome *f(x)* zadovoljava svojstva:
  1. $f(x) \geq 0, \quad \forall x$
  2. $\int_{-\infty}^{\infty} f(x)dx = 1$

Funkcija raspodjele

Funkcija raspodjele (kumulativna funkcija distribucije) diskretne slučajne varijable ***X*** definirana je:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy$$

- Općenito vrijedi:

  $P(a \leq X \leq b) = F(b) - F(a)$ → distribucijska funkcija ima važnu ulogu kod kontuiruiranih sl. var. jer pomoću nje jednostavno dobivamo vjerojatnost

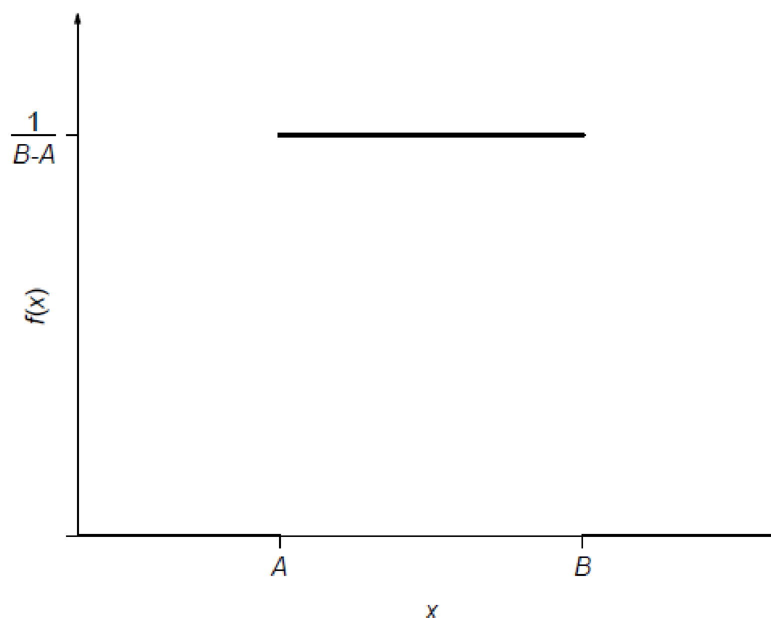
- Iz distribucijske funkcije jednostavno dobivamo funkciju gustoće vjerojatnosti iz:

$$f(x) = \frac{dF(x)}{dx}$$

Funkcija gustoće vjerojatnosti

Primjer: ***uniformna ili pravokutna raspodjela*** u intervalu [A,B]

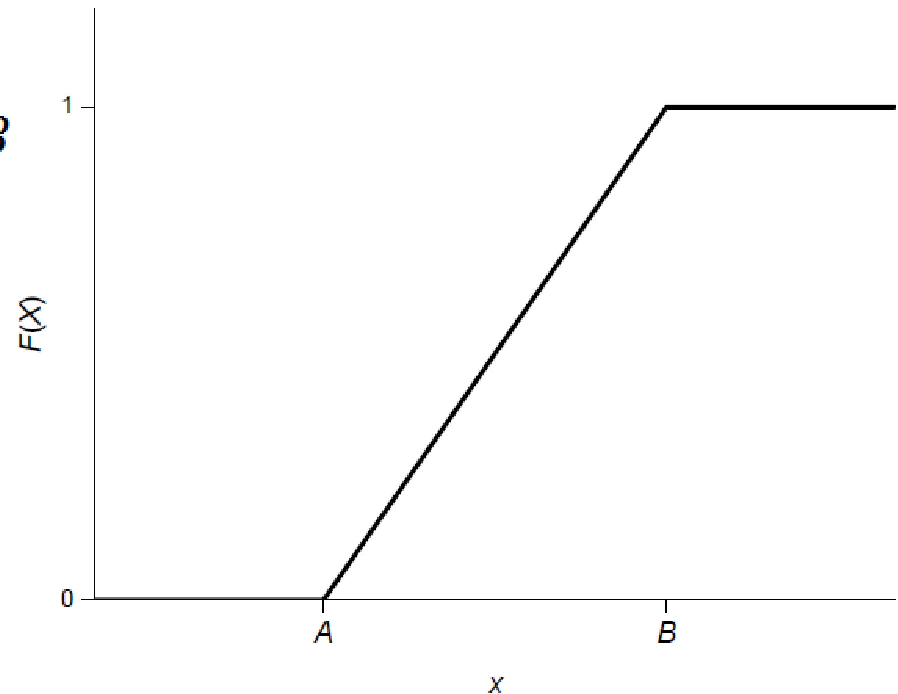
$$f(x) = \begin{cases} \dfrac{1}{B-A} & A \leq x \leq B \\ 0 & inače \end{cases}$$



Funkcija raspodjele

- Primjer: ***uniformna ili pravokutna raspodjela*** u intervalu [A,B]

$$F(x) = \begin{cases} 0, & x < A \\ \dfrac{x-A}{B-A}, & A \le x \le B \\ 1, & x > B \end{cases}$$



Očekivanje i varijanca kontinuirane slučajne varijable

Očekivanje kontinuirane slučajne varijable ***X*** definira se:

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Varijanca kontinuirane slučajne varijable ***X*** je:

$$V(X) = \sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = E[(X - \mu)^2]$$

Može se pokazati da i ovdje vrijedi:

$$V(X) = E(X^2) - [E(X)]^2$$

Očekivanje i varijanca kontinuirane slučajne varijable

- Primjer: ***uniformna ili pravokutna raspodjela*** u intervalu [A,B]

$$f(x) = \begin{cases} \dfrac{1}{B-A} & A \leq x \leq B \\ 0 & inače \end{cases}$$

- Očekivanje: $E(X) = \dfrac{A+B}{2}$

- Varijanca: $V(X) = \dfrac{(A-B)^2}{12}$

Funkcija izvodnica i momenti

Definicije:

- funkcija izvodnica:

$$g(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

- pomoćni moment:

$$m_r = \int_{-\infty}^{\infty} x^r f(x) dx$$

- centralni moment:

$$M_r = \int_{-\infty}^{\infty} (x - \mu)^r f(x) dx$$

Očekivanje i varijanca *funkcije* kontinuirane slučajne varijable

- Za ***funkciju h(X)*** kontinuirane slučajne varijable ***X*** s gustoćom vjerojatnosti ***f(x)***, očekivanje je:

$$E\big(h(X)\big) = \mu_{h(X)} = \int_{-\infty}^{\infty} h(x) f(x) dx$$

- Za ***funkciju h(X)*** kontinuirane slučajne varijable ***X*** s gustoćom vjerojatnosti ***f(x)***, varijanca je:

$$V\big(h(X)\big) = E[h(X)^2] - \big[E\big(h(X)\big)\big]^2$$

- <u>Specijalno</u>: linearna funkcija h(X)=aX+b
  - Očekivanje: **E(aX+b)=a E(X)+b**
  - Varijanca: **V(aX+b)=a^2 V(X)**

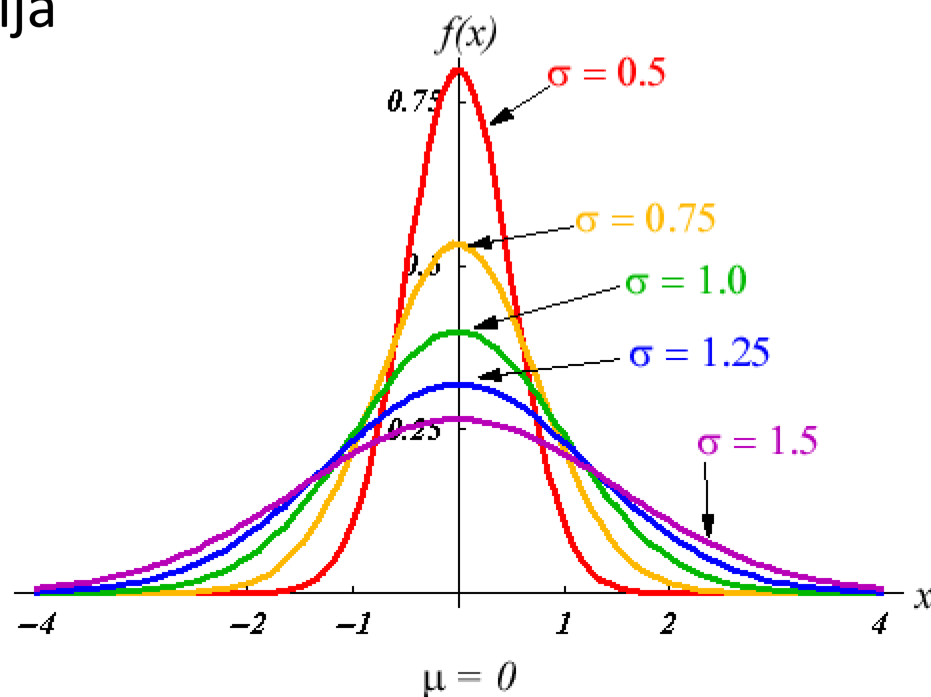
Normalna (Gaussova) raspodjela

- Slučajna varijabla ***X*** ima normalnu raspodjelu ako je njena funkcija gustoće vjerojatnosti:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

gdje x može biti iz $(-\infty, \infty)$

- Slučajnu varijablu koja ima normalnu raspodjelu označavamo sa: $X \sim N(\mu, \sigma^2)$



Normalna (Gaussova) raspodjela

Normiranost raspodjele slijedi iz definicije:

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

uz supstituciju: $u = x - \mu$

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{u^2}{2\sigma^2}} du = 2\frac{1}{\sigma\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{u^2}{2\sigma^2}} du = \frac{2}{\sigma\sqrt{2\pi}} \frac{\sqrt{\pi}}{\frac{2}{\sqrt{2}\sigma}} = 1$$

Slijedi: $\int_{-\infty}^{\infty} f(x)dx = 1$

<u>Iz tablica</u>:

$$\int_{0}^{\infty} e^{-a^2x^2} dx = \frac{\sqrt{\pi}}{2a}, a > 0$$

Normalna (Gaussova) raspodjela

Očekivanje – prema definiciji:

$$E(X) = \int_{-\infty}^{\infty} x\, f(x) dx = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

- Uz supstituciju: $u = \frac{x-\mu}{\sigma}$
- Slijedi: $E(X) = \mu$

Varijanca – prema definiciji se može pokazati da slijedi:

$$V(X) = \sigma^2$$

Normalna (Gaussova) raspodjela

Standardizirana raspodjela

- Uvodimo substituciju: $u = \frac{x-\mu}{\sigma}$
- Funkcija gustoće vjerojatnosti standardne normalne raspodjele:

$$\varphi(u) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{u^2}{2}}$$

→ Normalna raspodjela s očekivanjem 0 i varijancom 1 ***U~N(0,1)***

- Veza s "običnom" Gaussovom raspodjelom:

$$f(x) = \frac{1}{\sigma}\, \varphi(u)$$

- Kumulativna funkcija standardizirane normalne raspodjele:

$$\Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

→ Vrijednosti dane u tablicama!

Normalna (Gaussova) raspodjela

Vjerojatnost da x poprimi vrijednost u intervalu (x_1, x_2) je po definiciji:

$$P(x_1 < x < x_2) = \frac{1}{\sqrt{2\pi}\sigma} \int_{x_1}^{x_2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

- Prelaskom na standardiziranu raspodjelu, tj. uvođenjem supstitucije $u = \frac{x-\mu}{\sigma}$ vjerojatnost postaje:

$$P(x_1 < x < x_2) = \frac{1}{\sqrt{2\pi}} \int_{u_1}^{u_2} e^{-\frac{u^2}{2}} du$$

- Koristimo distribucijsku funkciju:

$$P(x_1 < x < x_2) = \Phi(u_2) - \Phi(u_1)$$

Normalna (Gaussova) raspodjela

Vjerojatnost – primjer

- Uzmimo normalnu raspodjelu μ=50 i σ=2
- Kolika je vjerojatnost da ***X*** poprimi vrijednost između 46 i 53

$$P(46 < X < 53) = ?$$

- Uz substituciju $u = \dfrac{x-\mu}{\sigma}$:

$$P(46 < X < 53) = P(-2 < U < 1{,}5)$$
$$= \Phi(1{,}5) - \Phi(-2)$$

- Tablice: http://www.stat.ufl.edu/~athienit/Tables/Ztable.pdf

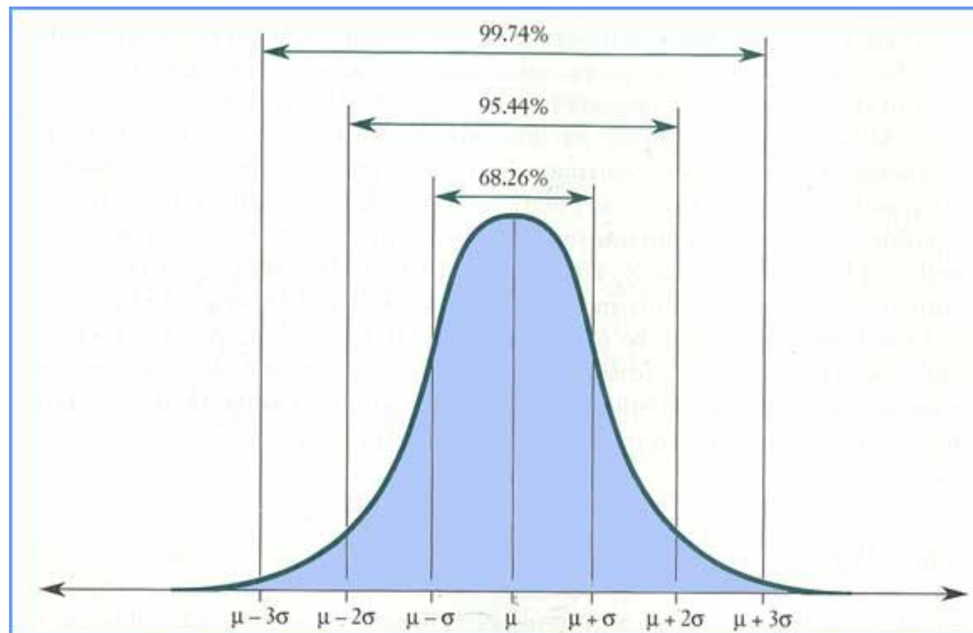
→ P = 0,9332-0,0228=0,9104

Normalna (Gaussova) raspodjela

- Vjerojatnost da X poprimi vrijednost:

Interval	P(x)
$(\mu-\sigma, \mu+\sigma)$	0,683
$(\mu-2\sigma, \mu+2\sigma)$	0,954
$(\mu-3\sigma, \mu+3\sigma)$	0,997

- Simetričnost $\alpha_3=0$
- Spljoštenost $\alpha_4=3$



Značaj normalne raspodjele

1. Objašnjava najveći broj statističih opažanja
2. Opisuje raspodjelu slučajne pogreške mjerenja
3. Općenito opisuje pojave gdje postoji velik broj slučajnih utjecaja → pokazuje sa da je ukupna raspodjela veličine podvrgnute tim utjecajima, Gaussova raspodjela

Aproksimacija binomne raspodjele Gaussovom

- Pokazuje se da kod binomne raspodjele:
  - koeficijent asimetrije $\alpha_3 \rightarrow 0$ kad $n \rightarrow \infty$ $\alpha_3 = \frac{q-p}{\sqrt{npq}}$
  - koeficijent spljoštenosti $\alpha_4 \rightarrow 3$ kad $n \rightarrow \infty$ $\alpha_4 = \frac{1-6pq}{npq} + 3$

  → to su upravo svojstva normalne raspodjele

  → Binomna teži u normalnu kad $n \rightarrow \infty$
- U praksi se uglavnom pokazuje da je aproksimacija Gaussovom raspodjelom dobra ako su zadovoljeni uvjeti:

1. $$\frac{1}{n+1} < p < \frac{n}{n+1}$$

2. $npq > 9$

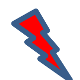
Aproksimacija binomne raspodjele Gaussovom

Primjer: koja je vjerojatnost da u 12 bacanja novčića ispadne najmanje 4 a najviše 7 glava?

- Binomna raspodjela:
  - ❑ n=12
  - ❑ p=1/2
  - → $P(4 \leq X \leq 7)$ = 0,733
- Aproksimacija Gaussovom:



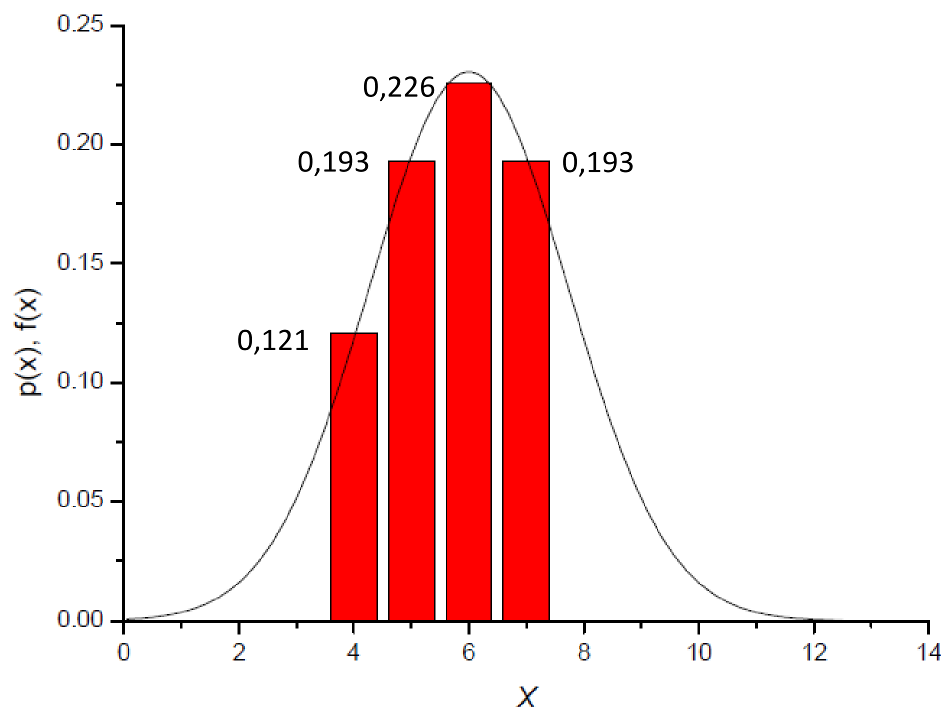
  - ➢ 1/13 < 1/2 < 12/13



  - ➢ npq=3<9

  μ = 6, σ^2 = 3

  → $P(3.5 < X < 7.5)$ = 0,733

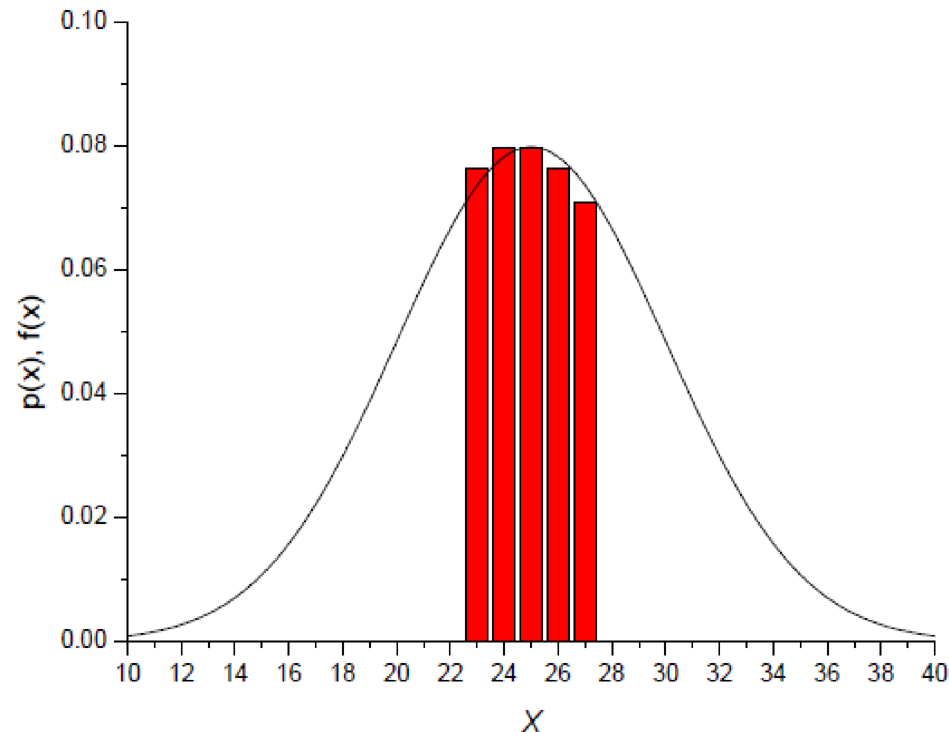


Aproksimacija Poissonove raspodjele Gaussovom

- Ako X ima Poissonovu raspodjelu X~P(λ) sa velikim λ (λ>20) može se dobro aproksimirati Gaussovom X~N(λ,λ)

<u>Primjer</u>: Geigerov brojač mjeri raspade sa srednjom brzinom od 25 s^{-1}

Koja je vjerojatnost da opazimo najmanje 23 i najviše 27 raspada/s ?



Poisson X~P(25):

$P(23 \leq X \leq 27)$ = 0,3826

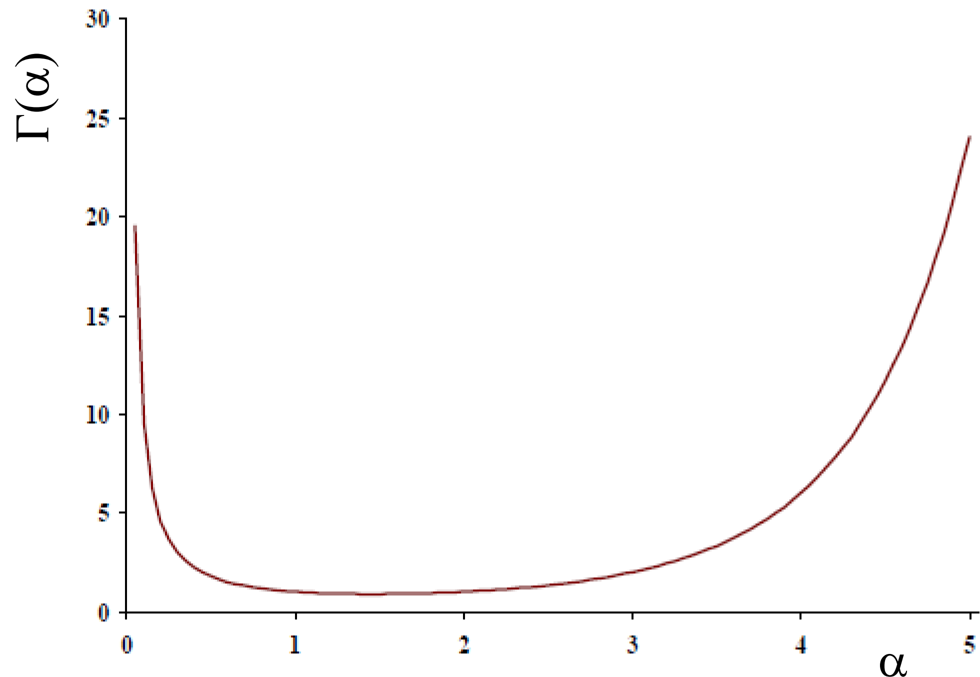
Gauss X~(25,25):

$P(22.5 < X < 27.5)$ = 0,3830

Obitelj gama raspodjela

Definiramo Γ-funkciju: $$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \qquad \alpha > 0$$

- Svojstva:
  1. $\forall \alpha > 1 \rightarrow \Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ (rekurzija)
  2. $\forall n \in \mathbb{N} \rightarrow \Gamma(n) = (n-1)!$ **(poopćenje faktorijela!)**
  3. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

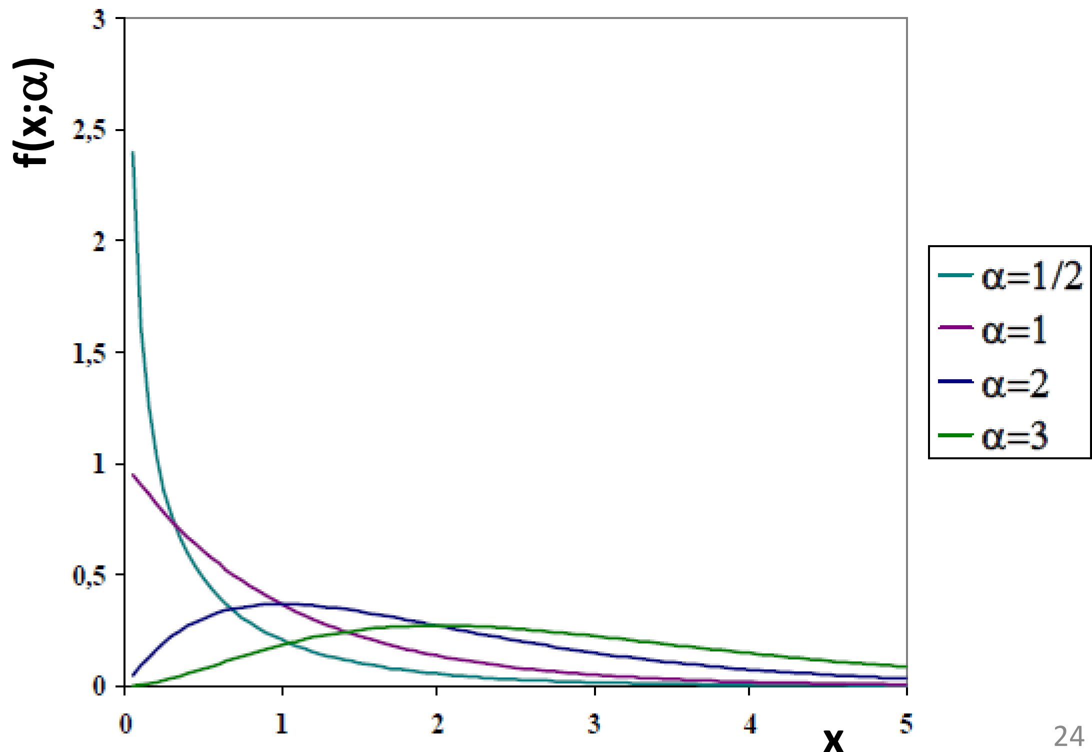


Standardna Γ-raspodjela

- Funkcija gustoće vjerojatnosti Γ raspodjele slučajne varijable ***X*** je:

$$f(x;\alpha) = \begin{cases} \dfrac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

- gdje je parametar raspodjele $\alpha>0$



Standardna Γ-raspodjela

- Svojstva – osnovni uvjeti raspodjele vjerojatnosti:

1. $f(x;\alpha) \geq 0, \qquad \forall x$
2. $\int_{-\infty}^{\infty} f(x;\alpha)dx = \frac{1}{\Gamma(\alpha)}\int_{0}^{\infty} x^{\alpha-1}e^{-x}dx = \frac{\Gamma(\alpha)}{\Gamma(\alpha)} = 1$

- Očekivanje: *E(X)*=α

- Varijanca: *V(X)*=α

Općenita Γ-raspodjela

- Funkcija gustoće vjerojatnosti:

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{x^{\alpha-1} e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

gdje su α, β >0

- Očekivanje: $E(X)=\alpha\beta$

- Varijanca: $V(X)=\alpha\beta^2$

Za β=1 općenita → standardna Γ-raspodjela

Statistika i osnovna mjerenja

Dvodimenzionalne raspodjele i korelacija

M. Makek

2016/2017

Višedimenzionalne raspodjele

Primjer: u eksperimentu sudaramo 2 protona

- Prije sudara oni imaju energiju i impuls E_1, p_1, E_2, p_2
- Nakon sudara nastaje mnoštvo različitih čestica s energijama i impulsima E_k',p_k'
- Da bismo odredili energiju i impuls čestice u konačnom stanju mjerimo npr. njen impuls u magnetskom spektrometru i energiju u kalorimetru.
- U tom slučaju energija i impuls su kontinuirane slučajne varijable
- Pitanja: da li su nezavisne? Ako ne, kako su one povezane? Kako su povezane sa energijama i impulsima ostalih čestica?

- U eksperimentima se često opaža veći broj varijabli, koje mogu biti neovisne ili povezane, o čemu će ovisiti analiza rezultata i račun pogreške

Dvodimenzionalne raspodjele

- Uzmimo **dvije** slučajne varijable **X** i **Y** definirane na istom prostoru elementranih događaja
- *Def*: **združena raspodjela vjerojatnosti** p(x,y) za diskretne slučajne varijable X i Y je vjerojatnost da istodobno X poprimi vrijednost x i Y poprimi vrijednost y:

$$p(x,y) = P(X=x \text{ i } Y=y)$$

- *Def*: Rubne raspodjele vjerojatnosti za varijable X i Y označavamo s $p_X(x)$ i $p_Y(y)$, a dane su izrazima:

$$p_X(x) = \sum_{y \in D_Y} p(x,y) \qquad p_Y(y) = \sum_{x \in D_X} p(x,y)$$

→ vjerojatnost da pojedina varijabla iz združene raspodjele poprimi neku vrijednost bez obzira na drugu varijablu.

Dvodimenzionalne raspodjele

- <u>Primjer</u>: studenti prve godine razvrstani po ocjeni iz statistike (varijabla X) i po spolu (varijabla Y)

Ocjena	Muško	Žensko	UKUPNO
1	10	2	12
2	13	3	16
3	15	7	22
4	18	7	25
5	4	3	7
UKUPNO	60	22	**82**

- *Vjerojatnost* da student bude žensko i ima ocjenu 3 je 7/82, a vjerojatnost student bude muško i ima ocjenu 3 je 15/82.
- *Rubna vjerojatnost* da bilo koji student ima ocjenu 3 je 22/82, a rubna vjerojatnost da student bude muško je 60/82

Dvodimenzionalne raspodjele

- Za kontinuirane slučajne varijable ***X,Y*** definiramo združenu funkciju gustoće vjerojatnosti *f(x,y)*, tako da za bilo koji dvodimenzionalni interval $A \subseteq \mathbb{R} \times \mathbb{R}$ vrijedi:

$$P((X,Y) \in A) = \iint_A f(x,y)dxdy$$

- Združena funkcija gustoće vjerojatnosti mora zadovoljavati uvjete: $f(x,y) \geq 0, \quad \forall x,y \quad \iint_{\mathbb{R}\times\mathbb{R}} f(x,y)dxdy = 1$

- Rubne funkcije gustoće vjerojatnosti definiramo kao:

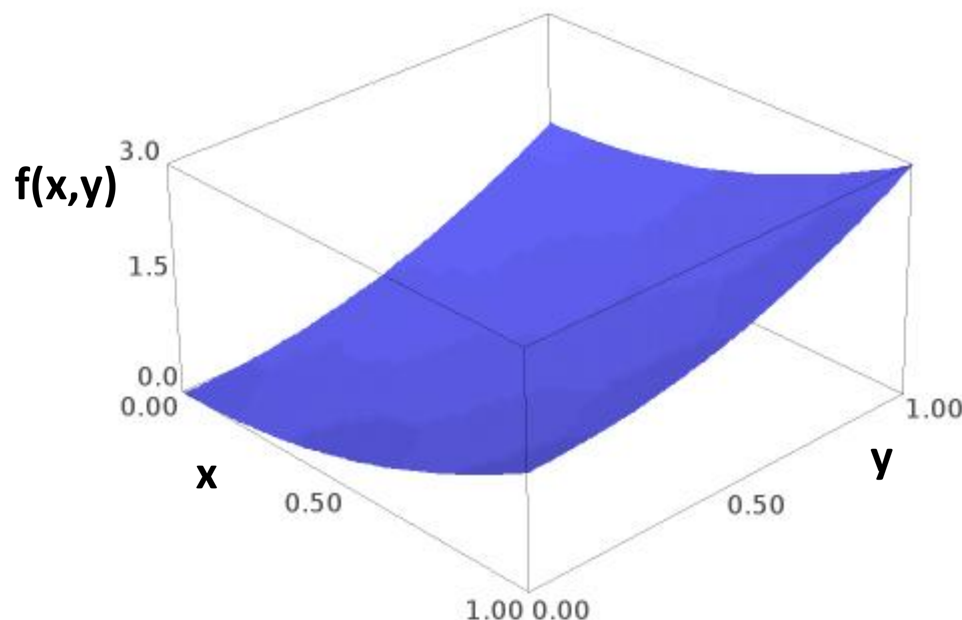
$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

Dvodimenzionalne raspodjele

- Primjer: združena funkcija gustoće vjerojatnosti je dana:

$$f(x,y)=\begin{cases}\frac{3}{2}(x^2+y^2), & x\in[0,1], y\in[0,1]\\ 0, & inače\end{cases}$$

- Koja je vjerojatnost da je X<1/4 i Y<1/4 ?
- Koja je vjerojatnost da je X< 1/4 bez obzira na Y ?



Uvjetna vjerojatnost

- Neka su ***X*** i ***Y*** kontinuirane slučajne varijable sa združenom funkcijom gustoće vjerojatnosti ***f(x,y)*** i rubnom funkcijom $f_X(x)$. Uvjetna funkcija gustoće vjerojatosti za varijablu ***Y***, ako je ***X=x*** je:

$$f_{(Y|X)} = \frac{f(x,y)}{f_X(x)}$$

Uz uvjet da je $f_X(x) \neq 0$

- Analogno pomoću raspodjela vjerojanosti definiramo uvjetnu vjerojatnost za diskretne varijable:

$$p_{(Y|X)} = \frac{p(x,y)}{p_X(x)}$$

Podsjetnik: nezavisnost događaja

- Definicija: događaji A i B su nezavisni onda i samo onda kad vrijedi:

  P (A|B) = P(A)

  (Vjerojatnost da se dogodi A ne ovisi o tome da li se dogodio B)

- Iz definicije uvjetne vjerojatnosti slijedi:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

- Iz toga slijedi:

$$\boxed{P(A \cap B) = P(A)P(B)}$$

Nezavisne slučajne varijable

- Svaka slučajna varijabla predstavlja ishod nekog događaja, stoga možemo poopćiti definiciju nezavisnih događaja na nezavisne varijable
- Dvije slučajne varijable će biti nezavisne, ako *vrijednost* jedne ne uvjetuje *vrijednost* druge, tj. ako vrijedi:

$$p_{(X|Y)} = \frac{p(x,y)}{p_Y(y)} = p_X(x)$$

- Iz toga slijedi da za nezavisne varijable vrijedi:

$$p(x,y) = p_X(x)p_Y(y)$$

Nezavisne slučajne varijable

- Općenito, slučajne varijable ***X*** i ***Y*** su nezavisne ako za svaki par vrijednosti x i y vrijedi:

$$p(x,y) = p_X(x)p_Y(y)$$ (za diskretne varijable)

$$f(x,y) = f_X(x)f_Y(y)$$ (za kontinuirane varijable)

- Na primjeru raspodjele studenata po ocjeni i po spolu vidimo da te varijable *nisu nezavisne*

Nezavisne slučajne varijable

- Primjer: istovremeno bacanje kocke i novčića
  - slučajna varijabla X opisuje ishod na kocki
  - Slučajna varijabla Y opisuju ishod na novčiću

Ishod	1	2	3	4	5	6	UKUPNO
P	1/12	1/12	1/12	1/12	1/12	1/12	**1/2**
G	1/12	1/12	1/12	1/12	1/12	1/12	**1/2**
UKUPNO	**1/6**	**1/6**	**1/6**	**1/6**	**1/6**	**1/6**	**1**

- Slijedi da za svaki par vrijednosti x i y vrijedi:

$$p(x,y) = p_X(x)p_Y(y)$$

- Dakle varijable su nezavisne!

Momenti dvodimenzionalne raspodjele

- Pomoćni moment:

$$m_{rs} = \sum_{x \in D_X} \sum_{y \in D_Y} x^r y^s p(x,y)$$ Za diskretne varijable

$$m_{rs} = \iint_{\mathbb{R}\times\mathbb{R}} x^r y^s f(x,y)\,dx\,dy$$ Za kontinuirane varijable

- Centralni moment:

$$M_{rs} = \sum_{x \in D_X} \sum_{y \in D_Y} (x-\mu_X)^r (y-\mu_Y)^s p(x,y)$$ Za diskretne varijable

$$M_{rs} = \iint_{\mathbb{R}\times\mathbb{R}} (x-\mu_X)^r (y-\mu_Y)^s f(x,y)\,dx\,dy$$ Za kontinuirane varijable

Očekivanje

- Pomoćni momenti $\boldsymbol{m_{01}}$ i $\boldsymbol{m_{10}}$:

$$m_{10} = \sum_{x \in D_X} \sum_{y \in D_Y} x\, p(x, y) = \sum_{x \in D_X} x p_X(x) = \mu_X$$

$$m_{01} = \sum_{x \in D_X} \sum_{y \in D_Y} y\, p(x, y) = \sum_{y \in D_Y} y p_Y(y) = \mu_Y$$

- Ovi pomoćni momenti odgovaraju **očekivanjima** diskretnih slučajnih varijabli ***X*** odn. ***Y***
- Analogno vrijedi i za kontinuirani slučaj

Varijanca

- Središnji momenti $\boldsymbol{M_{20}}$ i $\boldsymbol{M_{02}}$ odgovaraju varijancama:

$$M_{20} = \sum_{x \in D_X} \sum_{y \in D_Y} (x - \mu_X)^2 p(x, y)$$

$$= \sum_{x \in D_X} (x - \mu_X)^2 p_X(x) = V(X)$$

$$M_{02} = \sum_{x \in D_X} \sum_{y \in D_Y} (y - \mu_Y)^2 p(x, y)$$

$$= \sum_{y \in D_Y} (y - \mu_Y)^2 p_Y(y) = V(Y)$$

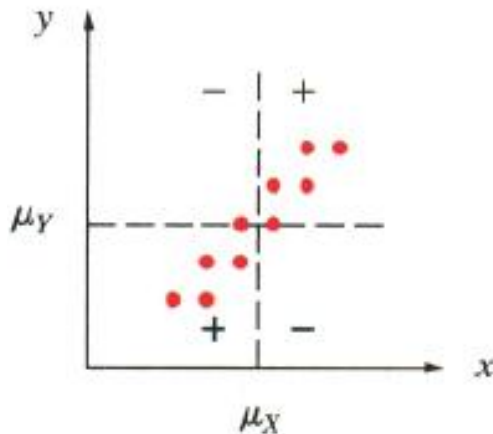
Kovarijanca

- Kovarijanca je veličina koja govori o zavisnosti varijabli ***X*** i ***Y***:

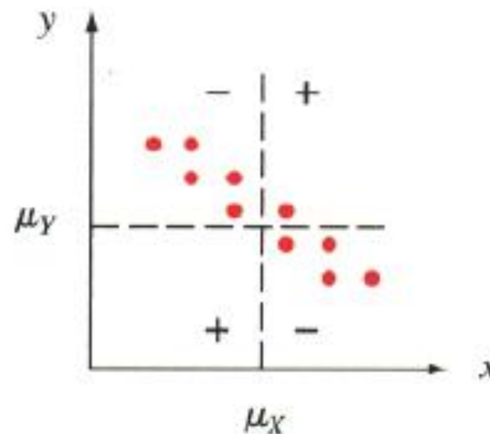
$$Cov(X,Y) = \sigma_{XY} = M_{11}$$

- Izražena pomoću pomoćnih momenata ova relacija postaje:

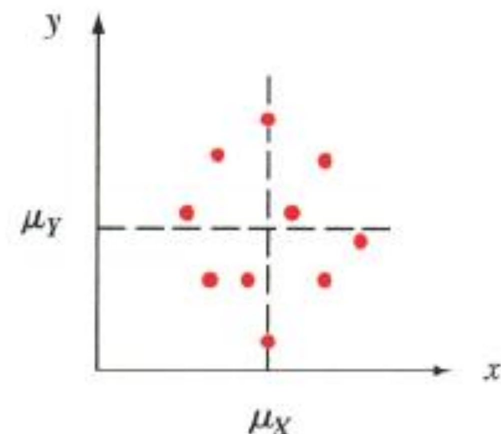
$$Cov(X,Y) = m_{11} - m_{10}m_{01} = E(XY) - E(X)E(Y)$$



Cov(X,Y) > 0



Cov(X,Y) < 0



Cov(X,Y) ~ 0

Kovarijanca nezavisnih varijabli

- Po definiciji za kovarijancu vrijedi:

$$\sigma_{xy} = E(XY) - E(X)E(Y)$$

- Prvi pribrojnik možemo pisati kao:

$$E(XY) = \sum_x \sum_y xy\, p(x, y)$$

- Za nezavisne varijable:

$$= \sum_x \sum_y xy\, p_x(x) p_y(y) = \sum_x x p_x(x) \sum_y y p_y(y)$$

- Slijedi: $E(XY) = E(X)E(Y)$
- Dobivamo da je kovarijanca nezavisnih varijabli: $\boxed{\sigma_{xy} = 0}$

Korelacija

- Kovarijanca ovisi o veličinama koje mjerimo i njihovim jedinicama → nije pogodno za usporedbu povezanosti različitih veličina
- Uvodimo bezdimenzionalnu veličinu – **korelaciju** – kao mjeru povezanosti, tj. zavisnosti varijabli
- **Koeficijent *linearne* korelacije** varijabli ***X*** i ***Y*** se definira kao:

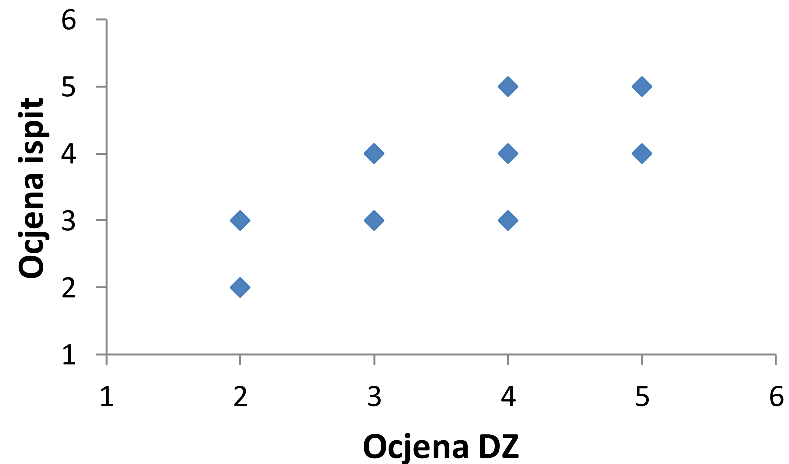
$$Corr(X,Y) = \rho_{X,Y} = \rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

(omjer kovarijance i produkta standardnih devijacija)

Korelacija

- <u>Primjer 1</u>: Za 10 studenata uspoređujemo ocjene iz domaće zadaće i ocjene na ispitu

Student	DZ	Ispit
1	5	5
2	4	5
3	4	4
4	2	3
5	5	4
6	2	2
7	3	4
8	3	3
9	4	3
10	3	4



$\sigma_{DZ} = 1{,}02$

$\sigma_{ISPIT} = 0{,}90$

$\sigma_{DZ\text{-}ISPIT} = 0{,}65$ (kovarijanca)

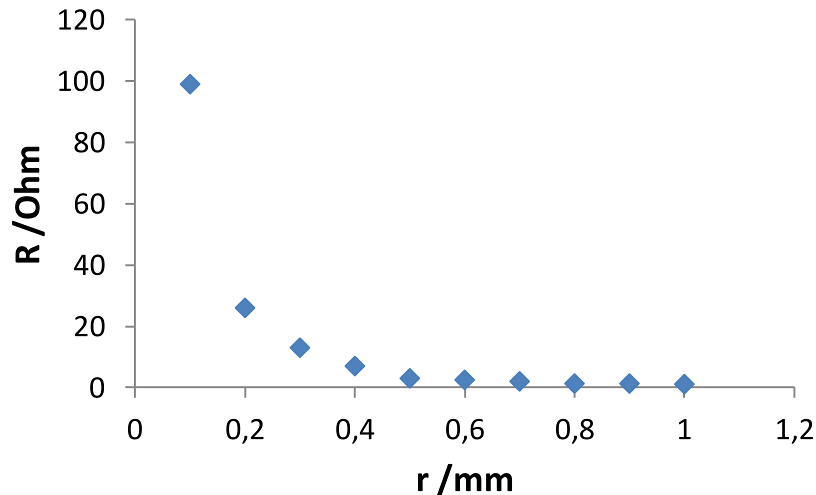
→ ρ = 0,7

Vidimo naznaku linearne ovisnosti

Korelacija

- Primjer 2: Studenti mjere otpor cilindričnog vodiča u ovisnosti o njegovom radijusu:

r /mm	R /Ω
0,1	99
0,2	26
0,3	13
0,4	7
0,5	3
0,6	2,5
0,7	2
0,8	1,30
0,9	1,30
1	1,10



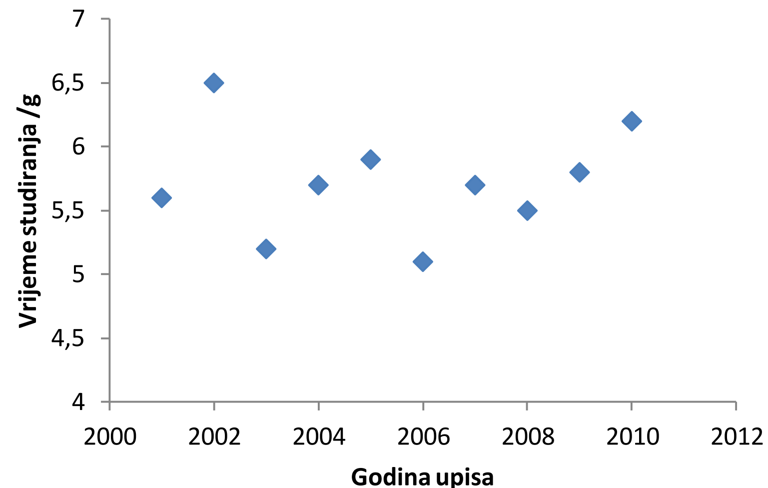
Koeficijent korelacije ρ = - 0,68

Vidimo izraženu obrnuto kvardatnu ovisnost, što teorijski i očekujemo: $R = \frac{1}{\sigma}\frac{l}{\pi r^2}$

Korelacija

- Primjer 3: promatramo vezu između godine upisa studija i trajanja studija

Godina	Trajanje /god
2001	5,6
2002	6,5
2003	5,2
2004	5,7
2005	5,9
2006	5,1
2007	5,7
2008	5,5
2009	5,8
2010	6,2



Koeficijent korelacije ρ = 0,05

Ne uočava se izražena ovisnost!

Korelacija

- Svojstva:
  1. $-1 \leq \rho \leq 1$
  2. Ako su ***X*** i ***Y*** nezavisne varijable onda je ρ=0 → primjer 3 (ali obrat ne vrijedi)
  3. Ako je ovisnost varijabli linearna Y=aX+b onda i samo onda je **ρ=1** ili **ρ=-1** (za a>0 odn. a<0)
  4. Može se još pokazati da vrijedi:

     $$Corr(aX + b, cY + d) = Corr(X, Y)$$

- Općenito ako je **|ρ|<0,5** kažemo da je linearna korelacija slaba (primjer 3), a ako je **|ρ|>0,5** kažemo da je linearna korelacija izražena (primjeri 1 i 2)
- Iz primjera 2 vidimo da imamo izrazito jaku kvadratnu korelaciju, no koeficijent je -0.68 obzirom da je on mjera linearne korelacija

Suma slučajnih varijabli

- Definiramo slučajnu varijablu $Y = X_1 + X_2$
- **Očekivanje** slučajne varijable ***Y***:

$$E(Y) = E(X_1) + E(X_2)$$

- Dokaz:

$$E(X_1 + X_2) = \sum_{x_1} \sum_{x_2} (x_1 + x_2) p(x_1, x_2)$$

$$= \sum_{x_1} \sum_{x_2} x_1 p(x_1, x_2) + \sum_{x_1} \sum_{x_2} x_2 p(x_1, x_2)$$

$$= \sum_{x_1} x_1 p_{x1}(x_1) + \sum_{x_2} x_2 p_{x2}(x_2)$$

$$= E(X_1) + E(X_2)$$

Suma slučajnih varijabli

- Definiramo slučajnu varijablu $Y=X_1+X_2$
- **Varijanca** slučajne varijable ***Y***:

$$V(Y) = V(X_1) + V(X_2)+2\sigma_{x1x2}$$

- Dokaz:

$$\begin{aligned} V(X_1 + X_2) &= E\{[(x_1 + x_2) - (\mu_1 + \mu_2)]^2\} \\ &= E\{[(x_1 - \mu_1) + (x_2 - \mu_2)]^2\} \\ &= E[(x_1 - \mu_1)^2] + E[(x_2 - \mu_2)^2] \\ &\quad + 2E[(x_1 - \mu_1)(x_2 - \mu_2)]) \end{aligned}$$

Linearna kombinacija slučajnih varijabli

- Za slučajne varijable X_1, X_2, ..., X_n možemo definirati novu slučajnu varijablu:

$$Y=a_1X_1+a_2X_2+\ldots+a_nX_n$$

gdje su $a_1 \ldots a_n$ konstatne. Y se naziva linearnom kombinacijom varijabli X.

- **Poopćenje relacije za očekivanje:**

$$E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n)$$

- Vrijedi bez obzira na to da li su varijable međusobno neovisne

Linearna kombinacija slučajnih varijabli

- **Poopćenje relacije za varijancu:**

$$V\left(a_1X_1 + a_2X_2 + \cdots + a_nX_n\right) = a_1^2V(X_1) + a_2^2V(X_2) + \cdots + a_n^2V(X_n) + \sum_i \sum_{j \neq i} a_i a_j Cov(X_i, X_j)$$

- Posebno, ako su varijable nezavisne, vrijedi:

$$V\left(a_1X_1 + a_2X_2 + \cdots + a_nX_n\right) = a_1^2V(X_1) + a_2^2V(X_2) + \cdots + a_n^2V(X_n)$$

- Važan primjer: **varijanca razlike nezavisnih varijabli**

$$V(X - Y) = V(X) + V(Y) = \sigma_x^2 + \sigma_y^2$$

Npr. mjerenje spektra s oduzimanjem pozadine.

Linearna kombinacija nezavisnih *normalnih* slučajnih varijabli

- Iz prethodnih razmatranja slijedi specijalni slučaj normalnih varijabli:

Neka su X_1, X_2,...,X_n nezavisne normalne slučajne varijable. Tada njihova linearna kombinacija

$$Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n$$

također ima normalnu raspodjelu s očekivanjem $\mu_Y = \sum_{i=1}^{n} a_i \mu_i$ i varijancom

$$\sigma_Y^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2 \; .$$

- Dokaz preskačemo

Statistika i osnovna mjerenja

Teorija uzoraka

M. Makek

2016/2017

Slučajni uzorak

- Iz neke populacije uzmemo uzorak od ***n*** – elemenata, koje označimo sa $x_1, \ldots, x_n$
- U mnogim statističkim problemima se elementi u uzorku mogu poistovjetiti sa opaženim vrijednostima ***n*** slučajnih varijabli $X_1, \ldots X_n$
- Definicija: ako su $X_1, \ldots X_n$ *slučajne varijable*, one čine ***slučajni uzorak*** veličine ***n*** ako su:
  - Nezavisne
  - Imaju iste raspodjele vjerojatnosti

  ($X_1, \ldots X_n$ su nezavisne i identično raspodijeljene)

Slučajni uzorak

- Svaku veličinu koju možemo izračunati iz uzorka nazivamo ***statistikom*** (npr. prosjek uzorka, standardna devijacija, momenti., itd.)
- Obzirom da prije uzimanja uzorka ne možemo znati vrijednost statistike ona je također **slučajna varijabla**
- Primjena:
  - Mjerenja fizikalnih veličina
  - Uzimanje uzorka iz velike populacije

Prosjek slučajnog uzorka

- Je definiran kao: $$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
- Prosjek je linearna kombinacija nezavisnih slučajnih varijabli $\boldsymbol{X_1, \ldots X_n}$ te je sam **slučajna varijabla**
- **Očekivanje i varijancu prosjeka** možemo odrediti prema relacijama za linearnu kombinaciju slučajnih varijabli:

$$E(\bar{X}) = E\left(\frac{1}{n}(X_1 + \cdots + X_n)\right)$$
$$= \frac{1}{n}\left(E(X_1) + \cdots + E(X_n)\right) = \frac{n\,\mu}{n} \quad \Rightarrow \quad \mu_{\bar{X}} = \mu$$

$$V(\bar{X}) = V\left(\frac{1}{n}(X_1 + \cdots + X_n)\right)$$
$$= \frac{1}{n^2}\left(V(X_1) + \cdots + V(X_n)\right) \quad \Rightarrow \quad \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$$

Total slučajnog uzorka

- Je definiran kao: $T_0 = \sum_{i=1}^{n} X_i$
- Total je linearna kombinacija nezavisnih slučajnih varijabli $\boldsymbol{X_1, \ldots X_n}$ te je sam **slučajna varijabla**
- **Očekivanje i varijancu totala** možemo odrediti prema relacijama za linearnu kombinaciju slučajnih varijabli:

$$E(T_0) = \mu_{T_0} = n\mu$$

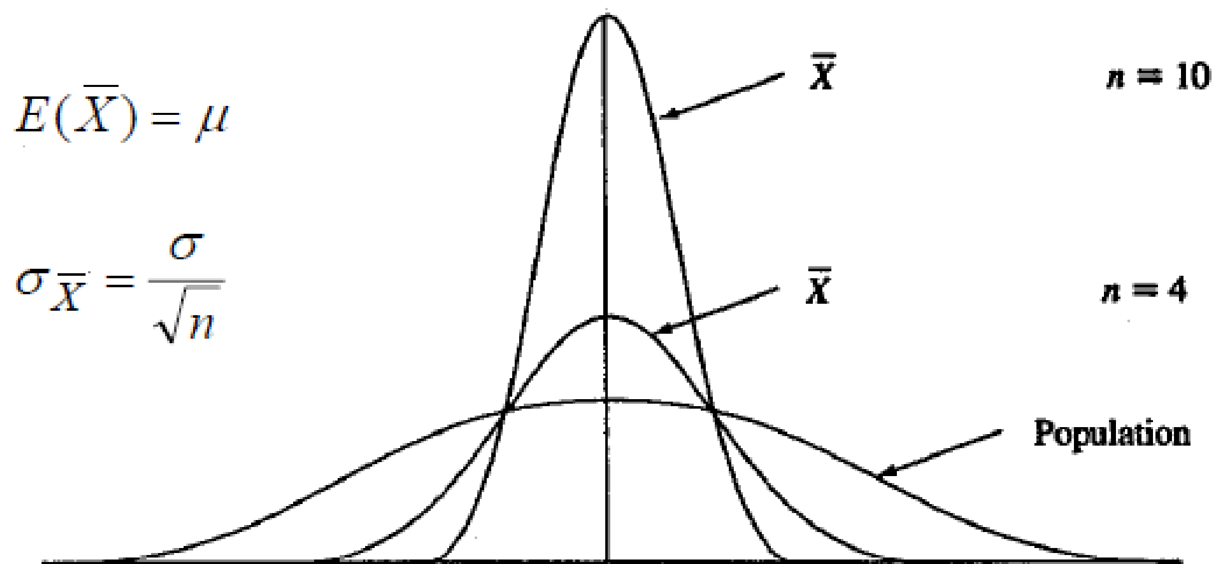
$$V(T_0) = \sigma_{T_0}^2 = n\sigma^2$$

Raspodjela prosjeka i totala

- Za slučajne varijable s **proizvoljnom raspodjelom** znamo ***očekivanje i varijancu*** prosjeka i totala no ne znamo točan oblik raspodjele
- Raspodjelu možemo dobiti iz *pravila vjerojatnosti* (za jedno-stavne raspodjele) ili pomoću *simulacijskog eksperimenta* (kad je nemoguće provesti analitički račun)

Raspodjela prosjeka normalne raspodjele

- Specijalni slučaj: normalna raspodjela
  - ako je uzorak sastavljen od ***n*** slučajnih varijabli s očekivanjem μ i varijancom σ^2, onda i prosjek ima normalnu raspodjelu s očekivanjem μ i varijancom σ^2/n (analogno vrijedi za total)
  - Što je ***n*** veći raspodjela prosjeka je uža (što imamo veći broj pokusa bolje je definirana srednja vrijednost)

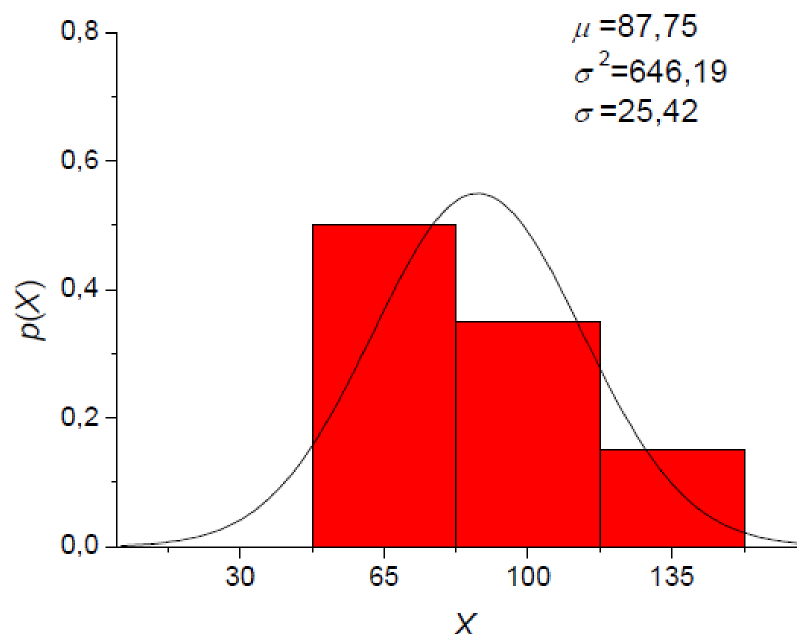


Primjer 1: diskretna raspodjela

Prodavač automobila prodaje 50% automobila niže klase po cijeni 65.000 kuna, 35% automobila srednje klase po 100.000 kuna i 15% automobila više klase po 135.000 kuna.
Definirajmo slučajnu varijablu

X=prihod od prodaje jednog automobila (u tisućama kuna)

x	$p(x)$
65	0,5
100	0,35
135	0,15



Primjer 1: diskretna raspodjela

Određenog dana najavila su se dva kupca. Neka su slučajne varijable:

X_1=prihod od prvog kupca

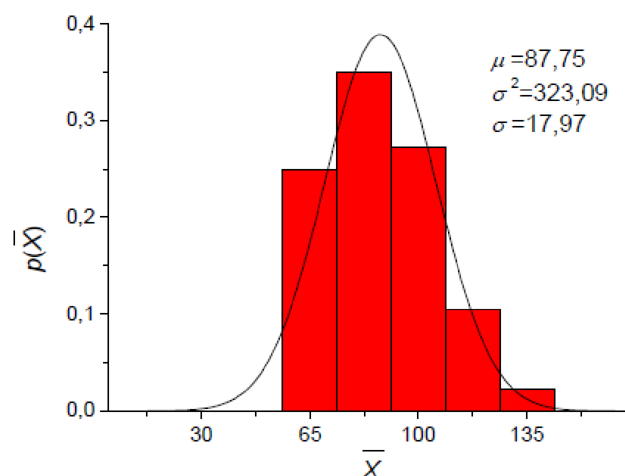
X_2=prihod od drugog kupca

Tablica mogućih prihoda:

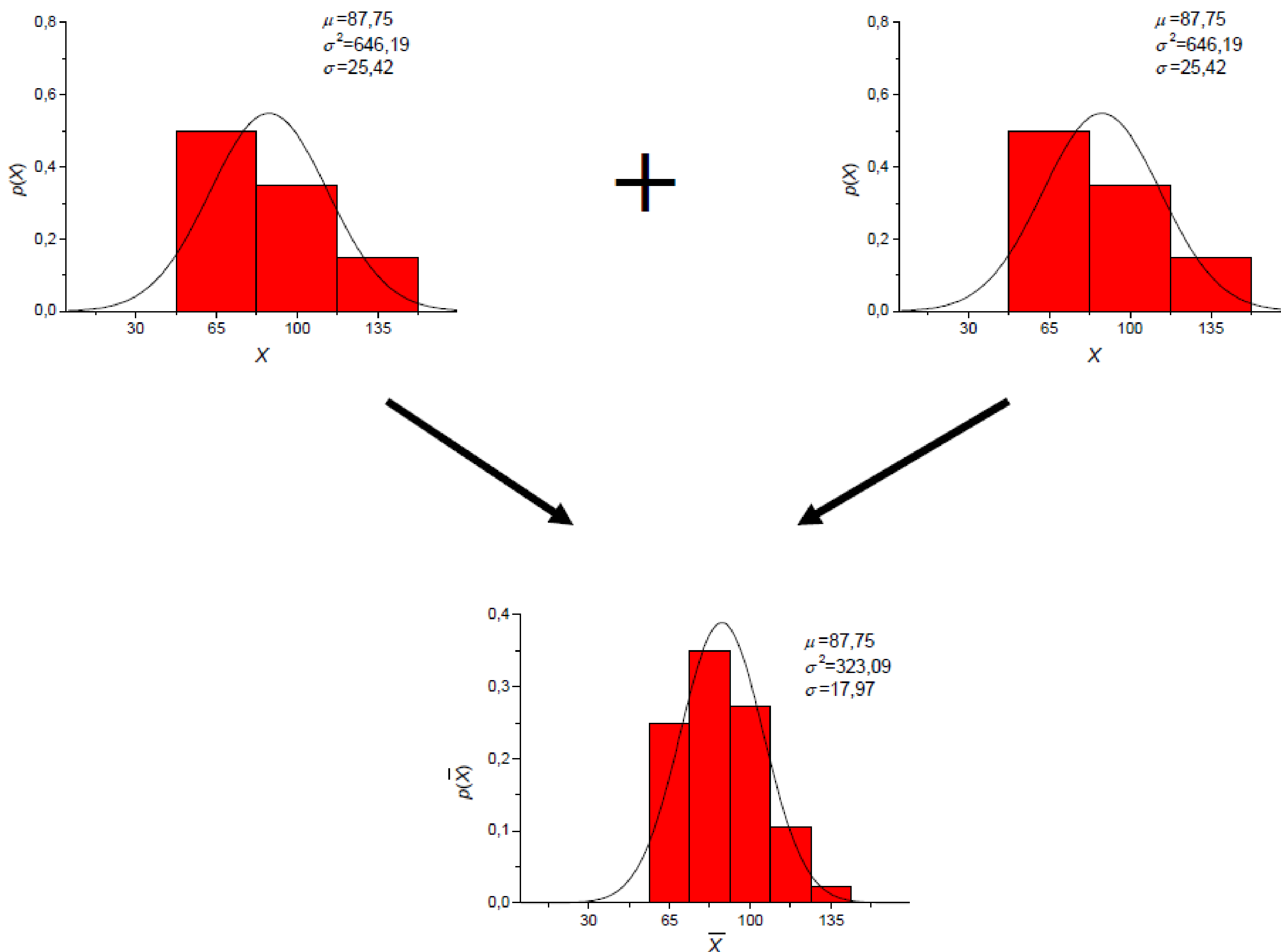
x_1	x_2	$p(x_1,x_2)$	x_1+x_2	$\bar{x}$
65	65	0,25	130	65
65	100	0,175	165	82,5
65	135	0,075	200	100
100	65	0,175	165	82,5
100	100	0,1225	200	100
100	135	0,0525	235	117,5
135	65	0,075	200	100
135	100	0,0525	235	117,5
135	135	0,0225	270	135

Slučajna varijabla $\overline{X}$ raspodijeljena je ovako:

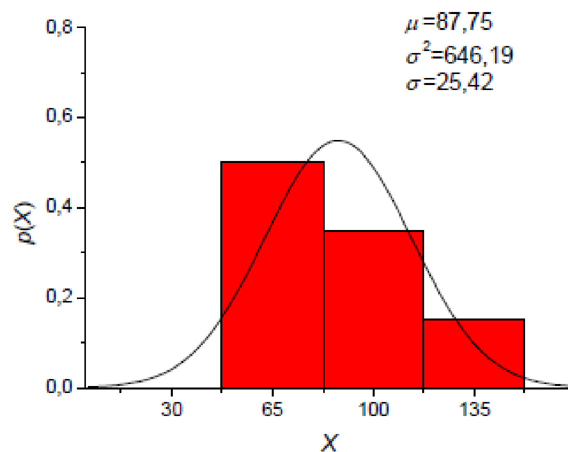
$\bar{x}$	$p(\bar{x})$
65	0,25
82,5	0,35
100	0,2725
117,5	0,105
135	0,0225

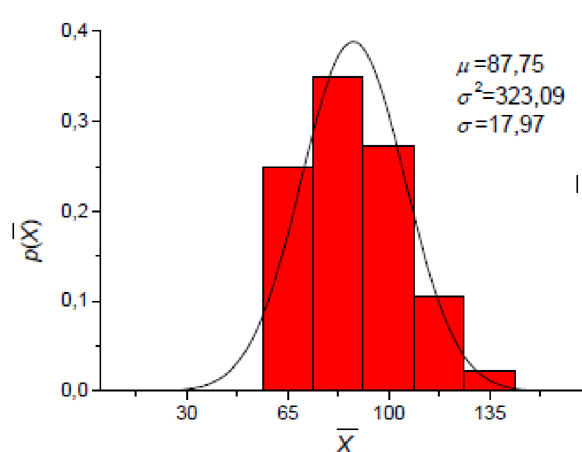


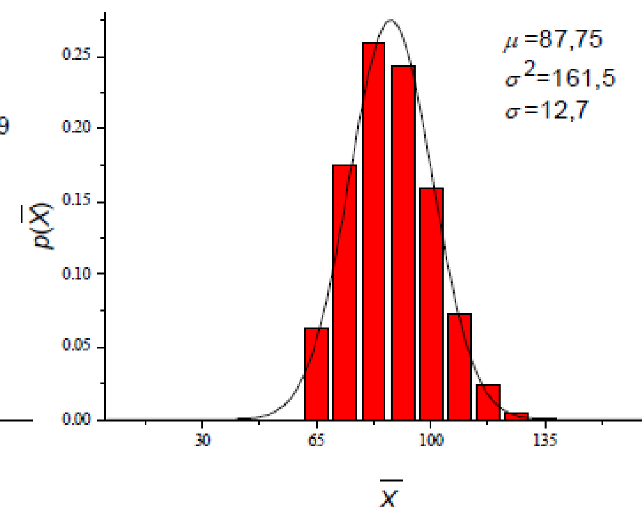
Primjer 1: diskretna raspodjela



Primjer 1: diskretna raspodjela







$n = 1$ $n = 2$ $n = 4$

Primjer 2: kontinuirana raspodjela

Na putu do posla čekam autobus koji vozi svakih 10 minuta, a zatim tramvaj koji također vozi svakih 10 minuta.

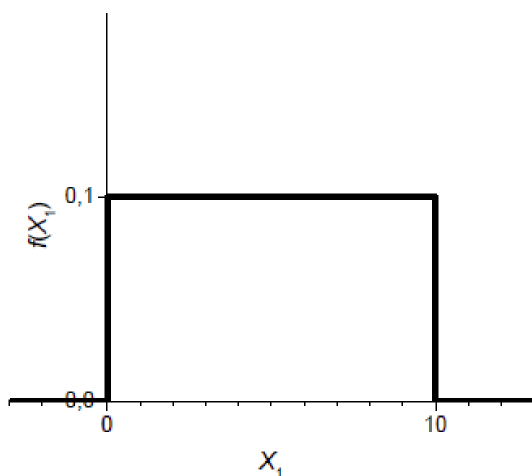
Neka su slučajne varijable:

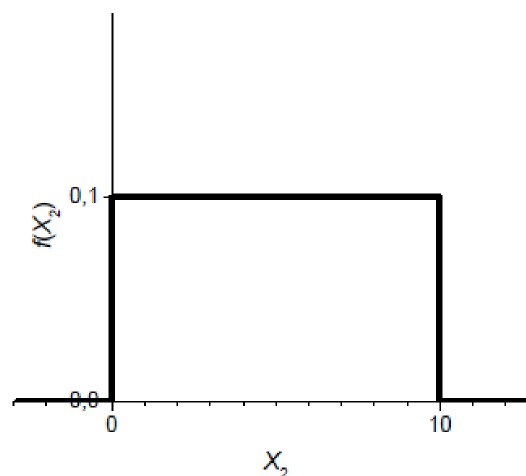
X_1 = vrijeme čekanja autobusa

X_2 = vrijeme čekanja tramvaja

$$f(x_1) = \begin{cases} 1/10 & , \quad 0 \le x_1 \le 10 \\ 0 & , \quad x_1 < 0 \quad ili \quad x_1 > 10 \end{cases}$$

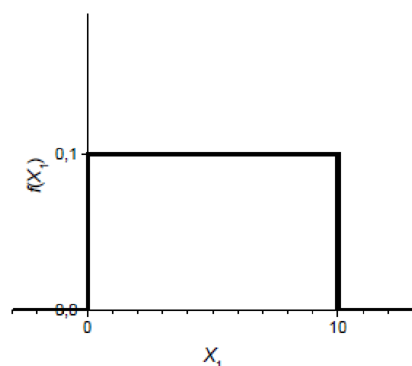
$$f(x_2) = \begin{cases} 1/10 & , \quad 0 \le x_2 \le 10 \\ 0 & , \quad x_2 < 0 \quad ili \quad x_2 > 10 \end{cases}$$



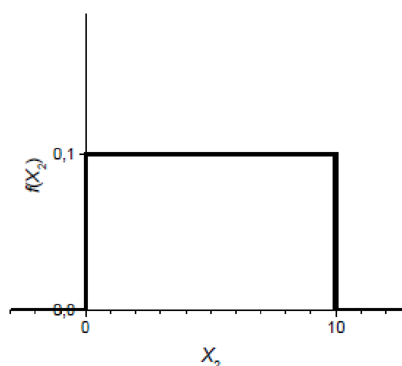


T_0 = ukupno vrijeme čekanja $= X_1 + X_2$

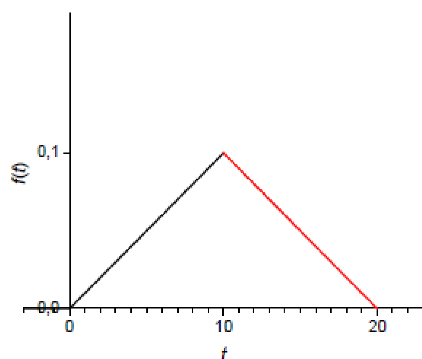
Primjer 2: kontinuirana raspodjela



+







$$f(t) = \begin{cases} \dfrac{t}{100}, 0 \leq t < 10 \\ \dfrac{20-t}{100}, 10 \leq t < 20 \\ 0, \quad inače \end{cases}$$

Uočiti

- Iz gornjih primjera vidi se da čak i kada populacija nije normalno raspodijeljena, ***raspodjela prosjeka i totala teži normalnoj raspodjeli*** kada je uzorak dovoljno velik

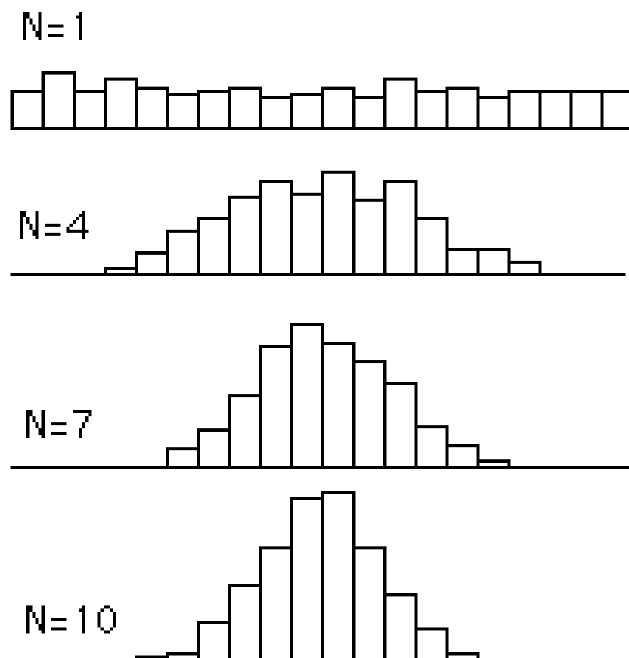
→ Centralni granični teorem

Centralni granični teorem

- Neka je $X_1, \dots X_n$ slučajni uzorak ***proizvoljne*** raspodjele očekivanjem μ i standardnom devijacijom σ. Ako je ***n*** dovoljno velik $\bar{X}$ ima približno ***normalnu*** raspodjelu s očekivanjem $\mu_{\bar{X}} = \mu$ i standardnom devijacijom $\sigma_{\bar{X}}^2 = \sigma^2/n$
  T_0 ima približno ***normalnu*** raspodjelu s očekivanjem $\mu_{T_0} = n\mu$ i standardnom devijacijom $\sigma_{T_0}^2 = n\sigma^2$
- U praksi se teorem može primijeniti za n>30

Centralni granični teorem

- Primjer:



- Simulacija:

http://onlinestatbook.com/stat_sim/sampling_dist/

Statistika i osnovna mjerenja

Procjenitelji i pogreške

M. Makek

2016/2017

Procjena parametara populacije

- Uzmimo da populacija ima karakteristične parametre poput μ, σ, σ^2, momenti, itd.
- Te parametre je nemoguće odrediti na cijeloj populaciji, stoga se uzima ***uzorak*** i na temelju njega se procjenjuju parametri populacije
- Uzmimo slučajnu varijablu Θ kao ***procjenitelj*** parametra θ
- Θ je nepristrani procjenitelj parametra θ ako vrijedi:

$$E(\Theta) = \theta$$

tj. mu je očekivanje stvarno jednako tom parametru

Procjena parametara populacije

- Ako je $X_1, \dots, X_n$ slučajni uzorak populacije, onda je slučajna varijabla $\bar{X}$ nepristrani procjenitelj očekivanja populacije μ

  Dokaz: pokazali smo da je očekivanje $\bar{X}$ jednako očekivanju populacije $E(\bar{X}) = \mu_{\bar{X}} = \mu$

- Ako je $X_1, \dots, X_n$ slučajni uzorak populacije, onda je slučajna varijabla S^2 nepristrani procjenitelj varijance populacije σ^2

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

- Dokaz slijedi

Procjena parametara populacije

- Pokazali smo da za bilo koju slučajnu varijablu ***Y*** vrijedi:

$$V(Y) = E(Y^2) - [E(Y)]^2$$

- Stoga za slučajnu varijablu $\boldsymbol{X_i}$ možemo pisati:

$$E\left(X_i^2\right) = [E(X_i)]^2 + V(X_i)$$

- Raspišemo izraz za $\boldsymbol{S^2}$:

$$S^2 = \frac{1}{n-1}\left[\sum X_i^2 - \frac{1}{n}\left(\sum X_i\right)^2\right]$$

Procjena parametara populacije

$$E(S^2) = \frac{1}{n-1}\left\{\sum E\left(X_i^2\right) - \frac{1}{n}E\left[\left(\sum X_i\right)^2\right]\right\}$$

$$= \frac{1}{n-1}\left\{\sum(\mu^2 + \sigma^2) - \frac{1}{n}\left\{V\left[\sum X_i\right] + \left[E\left(\sum X_i\right)\right]^2\right\}\right\}$$

$$= \frac{1}{n-1}\left\{\sum(\mu^2 + \sigma^2) - \frac{1}{n}\{V[T_0] + [E(T_0)]^2\}\right\}$$

$$= \frac{1}{n-1}\left\{\sum(\mu^2 + \sigma^2) - \frac{1}{n}\{n\sigma^2 + [n\mu]^2\}\right\} = \sigma^2$$

- Što je i trebalo pokazati

Intervali pouzdanosti

- Procjenom parametara dobivamo njihove iznose, no ne znamo koliko su ti iznosi pouzdani!
- Npr. ako izračunamo prosjek uzorka on neće biti identičan očekivanju populacije, no zanima nas koliko je njihovo odstupanje
- Prema središnjem graničnom teoremu, srednja vrijednost uzorka ima normalnu raspodjelu $\bar{X} \sim N(\mu, \sigma^2/n)$ gdje su $\boldsymbol{\mu}$ i $\boldsymbol{\sigma}$ očekivanje i varijanca populacije

Pouzdanost procjene prosjeka populacije

- Obzirom da srednja vrijednost uzorka ima normalnu raspodjelu, vrijedi:

$$P\left(-1 \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq 1\right) = 68\%$$

$$P\left(-1{,}96 \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq 1{,}96\right) = 95\%$$

$$P\left(-2{,}575 \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq 2{,}575\right) = 99\%$$

- Dakle možemo reći da se prava vrijednost očekivanja populacije nalazu i intervalu: $\bar{x} \pm \frac{\sigma}{\sqrt{n}}$

  s vjerojatnošću od 68%

Pouzdanost procjene prosjeka populacije

- u prethodnoj relaciji smo pretpostavili da je varijanca populacije σ poznata.
- Ako nije poznata onda je procjenjujemo pomoću slučajne varijable:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$

- Tada interval od 68% pouzdanosti postaje:

$$\bar{x} \pm \frac{s}{\sqrt{n}}$$

Pogreške mjerenja

- U fizici se rezultati mjerenja pišu kao srednja vrijednost nezavisnih mjerenja u intervalu od 68% pouzdanosti: $\bar{x} \pm M$
- Gdje je $M = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n(n-1)}}$ nepouzdanost prosjeka, koju nazivamo i ***standardnom greškom*** (sad znamo od kuda dolazi ova formula)
- Pri tome je rasipanje odn. Standardna devijacija uzorka: $m = s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$
- Obzirom da je rasipanje dolazi zbog nemogućnosti kontrole slučajnih pogrešaka, kažemo da je ***m*** mjera ***preciznosti aparature***
- ***m se ne mijenja značajno s povećanjem broja mjerenja, no M=m/√n se smanjuje te kažemo da je rezultat pouzdaniji ako napravimo više mjerenja***

Mjerenja različitih statističkih težina

- Mjerimo neku fizikalnu veličinu u nekoliko različitih mjerenja s različitim pouzdanostima → zanima nas procjena prave vrijednosti te veličine na temelju tih mjerenja
- Izvedeno je ***k*** nizova mjerenja te fizikalne veličine te je za svaki niz dobivena srednja vrijednost i standardna pogreška:

$$x_i = \bar{x}_i \pm M_i, \qquad i = 1 \dots k$$

- Tražimo najvjerojatniju vrijednost $\boldsymbol{x_p}$ mjerene fizikalne veličine na temelju tih mjerenja

Opća srednja vrijednost

- Funkcija gustoće vjerojatnosti je na temelju ***k*** serija mjerenja:

$$f\left(x_p; \bar{x}_1, \dots, \bar{x}_k\right) = \frac{1}{M_1 \dots M_k\left(\sqrt{2\pi}\right)^k} e^{-\sum_{i=1}^{k} \frac{\left(\bar{x}_i - x_p\right)^2}{2M_i^2}}$$

- Da bi našli za koji $\boldsymbol{x_p}$ je $\boldsymbol{f}$ maksimalan logaritmiramo i tražimo minimum sume:

$$\sum_{i=1}^{k} \frac{\left(\bar{x}_i - x_p\right)^2}{2M_i^2} = min.$$

Opća srednja vrijednost

- Iz uvjeta minimuma slijedi:

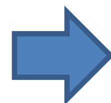
$$\frac{d}{dx_p}\sum_{i=1}^{k}\frac{\left(\bar{x}_i - x_p\right)^2}{2M_i^2} = -\sum_{i=1}^{k}\frac{\left(\bar{x}_i - x_p\right)}{M_i^2} = 0$$

$$x_p = \frac{\sum_{i=1}^{k}\frac{\bar{x}_i}{M_i^2}}{\sum_{i=1}^{k}\frac{1}{M_i^2}}$$

Pri tome definiramo statističke težine pojedine serije kao:

$$w_i = \frac{\frac{1}{M_i^2}}{\sum_{j=1}^{k}\frac{1}{M_j^2}}$$

- Slijedi opća srednja vrijednost:



$$x_p = \sum_{i=1}^{k} w_i \bar{x}_i$$

Nepouzdanost opće srednje vrijednosti

- Da bi odredili pouzdanost sjetimo se da je procjenitelj prave vrijednosti veličine, linearna kombinacija:

$$X_p = \sum_{i=1}^{k} w_i \bar{X}_i$$

- Prema relaciji za varijancu linearne kombinacije dobivamo:

$$M^2 = V\left(X_p\right) = \sum_{i=1}^{k} w_i^2 V(\bar{X}_i) = \sum_{i=1}^{k} w_i^2 M_i^2$$

- Primjetimo još da se statističke težine onose kao:

$$w_1 : w_2 : \ldots : w_n = \frac{1}{M_1^2} : \frac{1}{M_2^2} : \ldots : \frac{1}{M_n^2}$$

- I da vrijedi:

$$\sum_{i=1}^{k} w_i = 1$$

Propagacija pogreške

- Želimo odrediti neku fizikalnu veličinu preko njene funkcijske ovisnosti u neposredno mjerenim veličinama
- Primjeri:
  a) Mjerimo masu i volumen predmeta i želimo odrediti njegovu gustoću
  b) Mjerimo put i vrijeme koje auto prijeđe i želimo odrediti njegovu brzinu
- Pitanje je kako pogreška posredno mjerene veličine ovisi o pogreškama neposredno mjerenih veličina?
- Pri sljedećim razmatranjima pretpostavljamo da su pogreške mjerenja slučajne (normalno raspodijeljene) i malene tj. da vrijedi $M_i << \mu_i$
- Razmotrimo prvo dva specifična primjera koja ćemo poopćiti

Propagacija pogreške

1. Posredna veličina ***f*** je linearna kombinacija izravno mjerenih veličina ***X*** i ***Y***: ***f(X,Y)=aX+bY***
   -> Za slučaj linearne kombinacije *nezavisnih* varijabli znamo izračunati očekivanje i varijancu, dakle:
   $$E[f(X,Y)] = aE(X) + bE(Y) \qquad M_f = a^2 M_X^2 + b^2 M_Y^2$$
2. Posredna veličina ***f*** je nelinearna funkcija veličine ***X***: f(X)
   -> za očekivanje vrijedi (uz pretpostavku malene pogreške):
   $$E[f(X)] \simeq f(\mu_X) = f(\bar{X})$$
   -> za varijancu vrijedi:
   $$M_f^2 = V[f(X)] = E\left[\left(f(X) - f(\mu_X)\right)^2\right]$$

Propagacija pogreške

- dalje razvijemo f(X) u red: $f(X) = f(\mu_X) + (X - \mu_X) \left.\frac{df}{dX}\right|_{\mu_X} + \cdots$
- slijedi da je:

$$E\left[\left(f(X) - f(\mu_X)\right)^2\right] \simeq E\left[\left((X - \mu_X) \left.\frac{df}{dX}\right|_{\mu_X}\right)^2\right]$$

$$= \left(\left.\frac{df}{dX}\right|_{\mu_X}\right)^2 M_X^2$$

- odnosno: $f(X) - f(\mu_X) \simeq (X - \mu_X) \left.\frac{df}{dX}\right|_{\mu_X}$

- Dakle za male pogreške vrijedi: $M_f = \left|\left.\frac{df}{dX}\right|_{\mu_X}\right| M_X$

Propagacija pogreške

3. Općeniti slučaj – posredna veličina ***f*** je funkcija ***n*** izravno mjerenih neposrednih *nezavisnih* veličina $X_1, \dots, X_n$

   -> pokazuje se da tada vrijedi:

$$\bar{f}(X_1, \dots, X_n) = f(\bar{x}_1, \dots, \bar{x}_n)$$

$$M_f = \sqrt{\sum_{i=1}^{n} \left(\left. \frac{\partial f}{\partial X_i} \right|_{\bar{X}_1, \dots \bar{X}_n} \right)^2 M_{X_i}^2}$$

   To su relacije koje smo uveli na početku, a sada je objašnjeno odakle one dolaze

Propagacija pogrešaka s korelacijom

- Prethodne relacije za propagaciju pogreške vrijede kada su slučajne varijable ***X*** i ***Y*** nezavisne
- Sada promatramo općenit slučaja kada dopuštamo da mjerenja ***X*** i ***Y*** nisu međusobno nezavisna
- Uzmimo da smo izvršili ***n*** mjerenja ***X*** i ***Y*** i kao rezultat dobili (x_i, y_i), gdje $i=1 \dots n$. Na temelju dobivenih vrijednosti izračunamo ***n*** vrijednosti veličine $q_i = q(x_i, y_i)$
- Zanima nas srednja vrijednost i pogreška veličine ***q***

Propagacija pogrešaka s korelacijom

- Pretpostavljamo da su pogreške opservabli X i Y relativno male tj. da su odstupanja $x_i - \bar{x}$ i $y_i - \bar{y}$ mala tako da pri razvoju u red možemo uzeti samo linearne članove:

$$q_i \approx q(\bar{x},\bar{y}) + (x_i - \bar{x}) \left.\frac{\partial q}{\partial x}\right|_{\bar{x},\bar{y}} + (y_i - \bar{y}) \left.\frac{\partial q}{\partial y}\right|_{\bar{x},\bar{y}}$$

- Prema definiciji ***srednja vrijednost*** je:

$$\bar{q} = \frac{1}{n}\sum_{i=1}^{n} q_i = \frac{1}{n}\sum_{i=1}^{n}\left[q(\bar{x},\bar{y}) + \underbrace{\sum (x_i - \bar{x}) \left.\frac{\partial q}{\partial x}\right|_{\bar{x},\bar{y}}}_{0} + \underbrace{\sum (y_i - \bar{y}) \left.\frac{\partial q}{\partial y}\right|_{\bar{x},\bar{y}}}_{0}\right]$$

- Slijedi: $\boxed{\bar{q} = q(\bar{x},\bar{y})}$

Propagacija pogrešaka s korelacijom

- Prema definiciji standardnu devijaciju možemo dobiti prema:

$$\sigma_q^2 = \frac{1}{n}\sum_{i=1}^{n}(q_i - \bar{q})^2 = \frac{1}{n}\sum_{i=1}^{n}\left[(x_i - \bar{x})\frac{\partial q}{\partial x}\bigg|_{\bar{x},\bar{y}} + (y_i - \bar{y})\frac{\partial q}{\partial y}\bigg|_{\bar{x},\bar{y}}\right]^2$$

$$= \left(\frac{\partial q}{\partial x}\right)^2 \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 + \left(\frac{\partial q}{\partial y}\right)^2 \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 + 2\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$\Rightarrow \quad \sigma_q^2 = \left(\frac{\partial q}{\partial x}\bigg|_{\bar{x},\bar{y}}\right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y}\bigg|_{\bar{x},\bar{y}}\right)^2 \sigma_y^2 + 2\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\bigg|_{\bar{x},\bar{y}} \sigma_{xy}$$

- gdje je σ_{xy} kovarijanca
- Gornja relacija daje standardnu devijaciju veličine q bez obzira na to jesu li veličine X i Y povezane.
- Za nezavisne varijable X i Y kovarijanca je 0, a relacija za standardnu devijaciju se svodi na prva dva člana

Propagacija pogrešaka s korelacijom

- Ako kovarijanca ne iščezava kažemo da su pogreške opservabli ***X*** i ***Y*** korelirane, a standardna devijacija $\boldsymbol{\sigma_q}$ je očito različita nego u slučaju neovisnih, slučajnih pogrešaka.
- Može se pokazati da relacija za $\boldsymbol{\sigma_q}$ daje opću gornju granicu standardne devijacije veličine ***q***.
- Da bi to dokazali iskoristit ćemo ***Cauchy-Schwarzovu*** nejednakost: $\sigma_{xy} \leq \sigma_x \sigma_y$
- Uvrštavanjem te nejednakosti u izraz za standardnu devijaciju:

$$\sigma_q^2 \leq \left(\frac{\partial q}{\partial x}\bigg|_{\bar{x},\bar{y}}\right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y}\bigg|_{\bar{x},\bar{y}}\right)^2 \sigma_y^2 + 2\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\bigg|_{\bar{x},\bar{y}} \sigma_x \sigma_y$$

Propagacija pogrešaka s korelacijom

- Dalje slijedi: $$\sigma_q^2 \leq \left[\left|\frac{\partial q}{\partial x}\right|\sigma_x + \left|\frac{\partial q}{\partial y}\right|\sigma_y\right]^2$$

- Čime dobivamo gornju granicu za standardnu devijaciju bez obzira na korelaciju varijabli ***X*** i ***Y***

$$\sigma_q \leq \left|\frac{\partial q}{\partial x}\right|\sigma_x + \left|\frac{\partial q}{\partial y}\right|\sigma_y$$

Statistika i osnovna mjerenja

Testiranje hipoteza

M. Makek

2016/2017

Postavljanje hipoteza

- Postavljamo neku teoriju i želimo provjeriti da li je ona istinita
- Primjeri:
  1. Kocka (za bacanje) nije poštena
  2. Novi lijek je bolji od starog
  3. Teorijski model dobro opisuje izmjerene podatke
  4. Optuženi nije kriv
- Da bi se izvršila provjera postavljamo dvije hipoteze:
  a) H_0 – nul-hipoteza
  b) H_1 – alternativna hipoteza

  → Samo jedna od njih je točna!
- H_0 se smatra ispravnom dok se ne dokaže suprotno → H_0 i H_1 se ne tretiraju ravnopravno, već se konzervativnija postavlja kao istinita dok se ne dokaže suprotno.

Ako ne odbacimo H_0, to ne znači da je ona ispravna, nego samo da nemamo dovoljno dokaza da ju odbacimo.

Postavljanje problema

1. Prvo određujemo:
   - A - područje prihvaćanja hipoteze H_0
   - B - Područje odbacivanja hipoteze H_0 (kritično područje)
2. Uzimamo uzorak iz populacije koju istražujemo
3. Donosimo odluku:

		ODLUKA	
		Odbaci H_0	Prihvati H_0
ISTINA	H_0	Pogreška 1. vrste	Ispravan zaključak
	H_1	Ispravan zaključak	Pogreška 2. vrste

4. Određujemo vjerojatnost pogreške:
   - $\alpha = P(B|H_0)$, vjerojatnost da odbacimo H_0 kada je istinita
   - $\beta = P(A|H_1)$, vjerojatnost da prihvatimo H_0 kada je neistinita

   → Smatra se da je pogreška 1. vrste teža pogreška (npr. krivnja optuženika)

Signifikantnost testa

- Provjera (test) ima **signifikantnost** α, ako je vjerojatnost za pogrešku prve vrste **<** α. (Test razine α)
- Konvencionalno se za razinu signifikantnosti uzima:
  - α=0,05 (signifikantan)
  - α=0,01 (vrlo signifikantan)
- **Snaga testa** je vjerojatnost da test uputi na ispravnu odluku, tj. da odbacimo H_0 kad je uistinu neistinita: **P=1-**β

Primjer – bacanje novčića

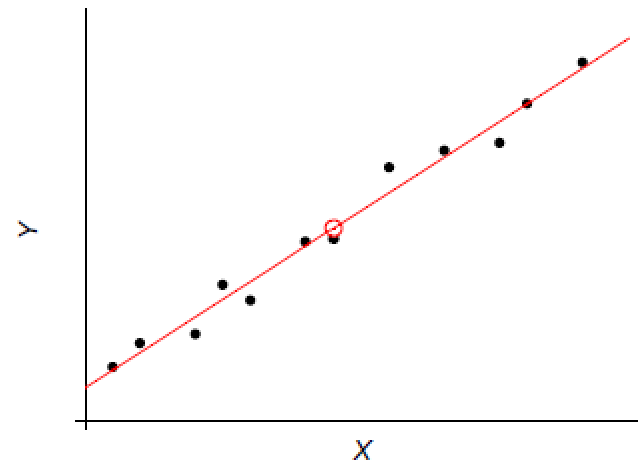
- Hipoteze:
  - H_0 = novčić je pošten
  - H_1 = novčić nije pošten
- X = broj 'pisama' u 6 bacanja → ako je novčić pošten onda je vjerojatnost $P(X|H_0) \sim Bin(6; 0.5)$
- Mogući su sljedeći ishodi:

X	0	1	2	3	4	5	6
$P(X\|H_0)$	0,0156	0,0938	0,2344	0,3125	0,2344	0,0938	0,0156

- Odaberemo: **A** za X=1,2,3,4,5 i **B** za X=0,6
- Vjerojatnost pogreške 1. vrste $\boldsymbol{\alpha=P(B|H_0)}$=0,0156 x 2
- Dakle ako u 6 bacanja kocke dobijemo 0 ili 6 puta pismo i na temelju toga zaključimo da je novčić nepošten, vjerojatnost da smo donjeli pogrešnu odluku je 3,12%

Prilagodba krivulje mjerenim podacima

- Pretpostavimo da smo izmjerili neku raspodjelu događaja (raspodjelu frekvencija)
- Prema obliku raspodjele i teoriji (ako je poznata) možemo mjerenim podacima prilagoditi krivulju regre-sijskim metodama (pravac, polinom, Gaussijan, Poissonova raspodijela, binomna raspodjela, itd.)



- Obzirom da svako mjerenje ima ograničenu preciznost javit će se odstupanja od krivulje
- Pitamo se da li su ta odstupanja slučajne prirode i da li pretpostavljena krivulja dobro opisuje izmjerene podatke

Dobrota prilagodbe

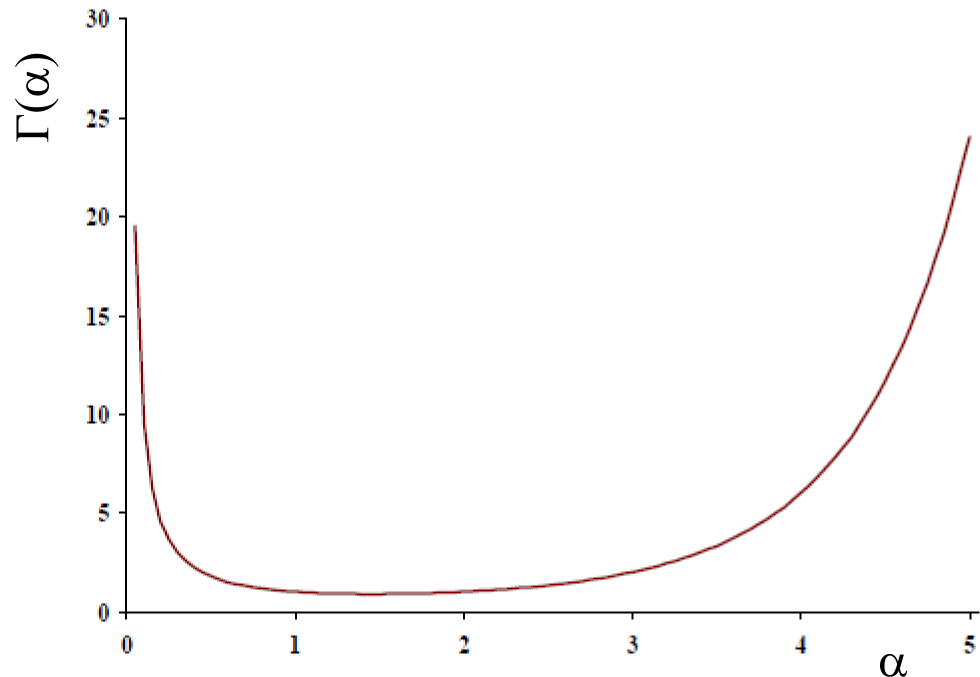
- Da bi ispitali kvalitetu (dobrotu) prilagodbe izmjerenim podacima izvodimo statistički test
- Jedan od testova koji se najčešće koriste je **χ^2-test**
- Podsjetimo se $\Gamma-$ funkcije, općenite $\Gamma-$ raspodjele i specifično χ^2 raspodjele

Gama funkcija

Definiramo Γ-funkciju:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \qquad \alpha > 0$$

- Svojstva:
  1. $\forall \alpha > 1 \rightarrow \Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ (rekurzija)
  2. $\forall n \in \mathbb{N} \rightarrow \Gamma(n) = (n-1)!$ **(poopćenje faktorijela!)**
  3. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$



Općenita Γ-raspodjela

- Funkcija gustoće vjerojatnosti:

$$f(x;\alpha,\beta)=\begin{cases}\dfrac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}, & x\geq 0\\ 0, & x<0\end{cases}$$

gdje su α, $\beta > 0$

- Očekivanje: $E(X)=\alpha\beta$
- Varijanca: $V(X)=\alpha\beta^2$

χ^2-raspodjela

- Specijalni slučaj općenite Γ raspodjele za β=2 i α=ν/2
- Funkcija gustoće vjerojatnosti:

$$f(x;\nu) = \begin{cases} \dfrac{x^{\nu/2-1}e^{-x/2}}{2^{\nu/2}\Gamma(\nu/2)}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

gdje je ν prirodan broj i naziva se “broj stupnjeva slobode”

- Veza s normalnom raspodjelom:

  Tm. Ako je ***u*** varijabla koja ima normalnu raspodjelu s očekivanjem i standardnom devijacijom μ i σ, onda je varijabla

$$x = \left(\frac{u-\mu}{\sigma}\right)^2$$

  raspodijeljena po χ^2-raspodjeli. Dokaz preskačemo.

χ^2-test

1. Uzmimo da eksperiment daje raspodjelu frekvencija f_i, koje grupiramo u ***n*** razreda (i=1…n).
2. Izmjerenim podacima pomoću regresije prilagodimo krivulju, koja daje teorijske frekvencije za pojedini razred f_{ti}
3. Definiramo veličinu:

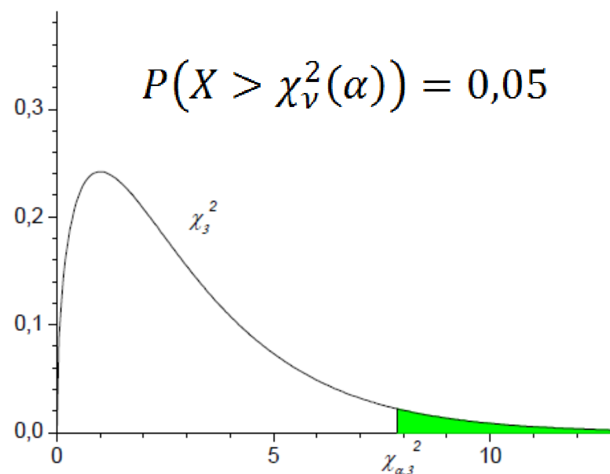
$$\chi^2 = \sum_{i=1}^{n} \frac{(f_i - f_{ti})^2}{f_{ti}}$$

4. ***Paersonov teorem***: veličina χ^2 je približno raspodijeljena prema χ^2-raspodjeli sa ν stupnjeva slobode koji ovise o broju opažanja (**n**) i broju ograničenja (**k**): $\nu = n - k$
5. Pri tome diskretnu raspodjelu aproksimiramo kontinuiranom. Aproksimacija ne vrijedi ako je $f_{ti} < 5$ → u tom slučaju treba grupirati razrede tako da je frekvencija pojedinog razreda uvijek veća od 4

Signifikantnost χ^2-testa

- Ako su razlike $|f_i - f_{ti}|$ relativno male, možemo zaključiti da imaju slučajan karakter
- Ako su razlike $|f_i - f_{ti}|$ relativno velike (signifikantne), onda će i χ^2 biti velik (signifikantan), tj. možemo zaključiti da je prevelik da bi bio slučajan. Tada je razumno odbaciti hipotezu.
- Vjerojatnost da varijabla ***X*** poprimi vrijednost veću od neke zadane $\boldsymbol{\chi_\nu^2(\alpha)}$, jednaka je površini repa ispod krivulje i dana je integralom:

$$P\left(X > \chi_\nu^2(\alpha)\right) = \int_{\chi_\nu^2(\alpha)}^{\infty} f(x;\nu)\, dx$$



- Po dogovoru se uzima da je χ^2 signifikantan ako padne u kritičnih 5% raspodjele i u tom slučaju odbacujemo hipotezu
- U prosječno 5% slučajeva će odbacivanje biti pogrešno jer je u toliko slučajeva moguće da je χ^2 signifikantan, a hipoteza je istinita.

Primjer 1

- Neki uređaj daje određeni postotak neispravnih proizvoda. Proizvodi se pakiraju u kutije od 20 komada. Za probu se otvara 50 kutija i određuje se broj neispravnih proizvoda X
- Opaženo je sljedeće:

X	0	1	2	3	>=4
f(x)	5	12	15	11	7

→ ***srednji x=1,8***

- ***Koja je vjerojatnost da se u kutiji nađe X defektnih komada?***
  - teorijski bi trebala biti dana binomnom raspodjelom: za $\bar{x} = np$ i n=20, dobivamo da je p=0.09

$$b(x) = \binom{20}{x} p^x q^{20-x}$$

  - Teorijske frekvencije:

X	0	1	2	3	>=4
$f_t(x)$	7,6	15,0	14,1	8,3	5,0

Primjer 1

- Nul-hipoteza: binomna raspodjela uz pretpostavku od 9% defektnih proizvoda dobro opisuje podatke
- Želimo potvrditi/odbaciti $\mathbf{H_0}$ uz signifikantnost **0.05**
- Broj stupnjeva slobode:
  - Broj razreda n=5
  - Broj ograničenja: 2 (određen **p** i **N**=50)
  - ν=5-2=3
- Izračunamo $\chi^2 = 3,08$
- Iz tablica vidimo da je za 3 stupnja slobode i 5% signifikantnosti $\chi^2_{0,05;3}$=7,82
- Obzirom da dobivena vrijednost ne spada u kritično područje tj. $\boldsymbol{\chi^2 < \chi^2_{0,05;3}}$ zadržavamo nul hipotezu.

Primjer 2

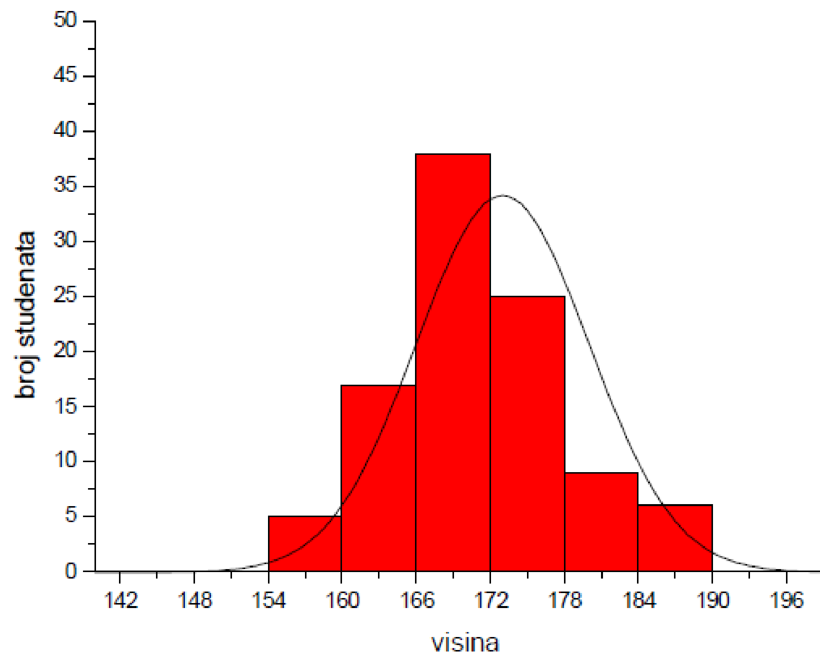
Dugogodišnje statistike pokazuju da je visina studenata na nekom sveučilištu normalno raspodijeljena s očekivanjem $\mu = 173$ cm i standardnom devijacijom $\sigma = 7$ cm.
Iz jedne generacije studenata na tom sveučilištu izdvojeno je 100 studenata, mjerene su njihove visine i svrstane u razrede.
Opažene su frekvencije:

x	154-160	160-166	166-172	172-178	178-184	184-190	Total
$f(x)$	5	17	38	25	9	6	100

Provjeri s 5% signifikantnosti je li raspodjela normalna s s očekivanjem $\mu = 173$ cm i standardnom devijacijom $\sigma = 7$ cm !

Primjer 2

H_0: *Raspodjela je Gaussova s očekivanjem* $\mu = 173$ cm *i standardnom devijacijom* $\sigma = 7$ cm.



Imamo jedno ograničenje: $\Sigma f_i = \Sigma f_{ti} = N = 100$

Primjer 2

- Izmjerene i teorijske frekvencije su dane tablicom:

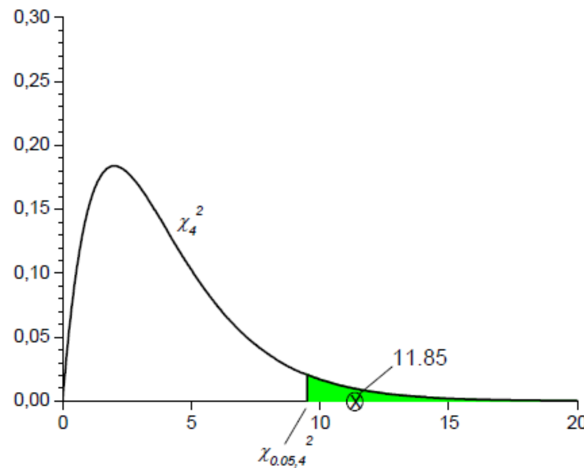
razred i	f_i	f_{ti}	$(f_i-f_{ti})^2/f_{ti}$
154-160	5	3	
160-166	17	13	2,25
166-172	38	28	3,57
172-178	25	32	1,53
178-184	9	18	4,5
184-190	6	6	0
zbroj	100	100	11,85

- Tablica ima 5 razreda te imamo jedno ograničenje N=100 →broj stupnjeva slobode je 4

Primjer 2

Za 4 stupnja slobode i 5% signifikantnosti, kritična vrijednost (iz tablica) je

$$\chi^2_{0.05,4} = 9.49$$



Izračunata vrijednost $\chi^2_{op} = 11.85$ pada u kritično područje.

Stoga hipotezu H_0 **odbacujemo**.

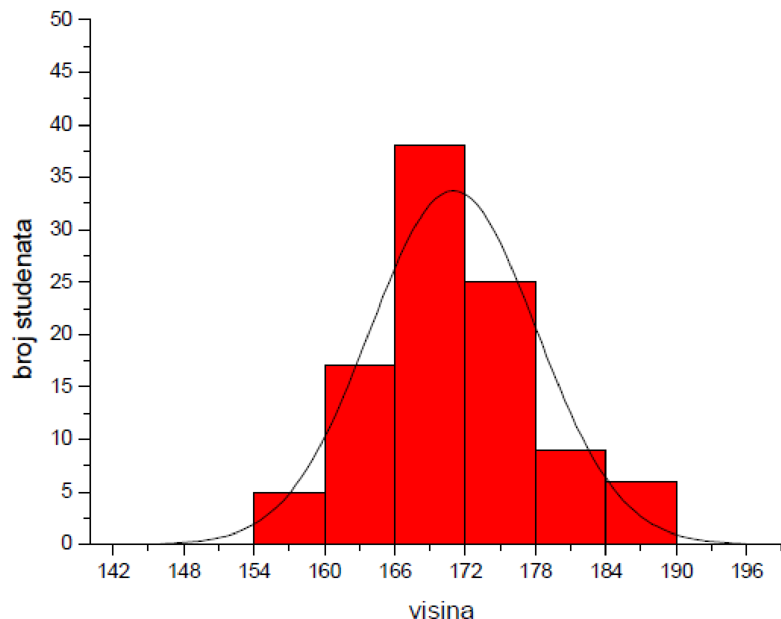
Primjer 2

- Nova hipoteza:

H$_0$: *Raspodjela je Gaussova s očekivanjem* $\mu = \bar{x}_{\text{uzorka}}$ *i standardnom devijacijom* $\sigma = \sigma_{\text{uzorka}}$.

Izračunamo:

$$\bar{x}_{\text{uzorka}} = \frac{1}{N}\sum f_i x_i = 171\,\text{cm} \qquad \sigma_{\text{uzorka}} = \sqrt{\frac{1}{N}\sum f_i (x_i - \bar{x}_{\text{uzorka}})^2} = 7{,}1\ \text{cm}$$



Primjer 2

- Izmjerene i teorijske frekvencije su dane tablicom:

razred *i*	f_i	f_{ti}	$(f_i-f_{ti})^2/f_{ti}$
154-160	5	6	0,167
160-166	17	18	0,056
166-172	38	32	1,125
172-178	25	28	0,321
178-184	9	13	0,063
184-190	6	3	
zbroj	100	100	1,732

Dakle, imamo pet razreda i tri ograničenja: $N = 100$

$$\mu = \bar{x}_{\text{uzorka}}$$

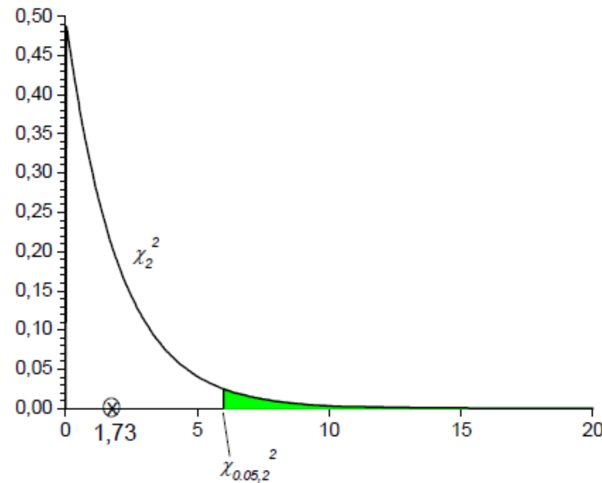
$$\sigma = \sigma_{\text{uzorka}}$$

pa je $\nu = 5\text{-}3 = 2$

Primjer 2

Za 2 stupnja slobode i 5% signifikantnosti, kritična vrijednost je

$$\chi^2_{0.05,2} = 5{,}99$$



Izračunata vrijednost $\chi^2_{op} = 1{,}73$ ne pada u kritično područje.
Stoga hipotezu H_0 **zadržavamo**.

Dodatak: χ^2-tablica

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997